

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: July 11, 2002, 08:14:43 ; Search time 5416.12 seconds
(without alignments)
8658.665 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 2241
Sequence: 1 tatectgtgggttgcccg.....aaaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.in.*
32: em.htg.other.*
33: em.htg.in.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

RESULT 1
MMU67176
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES

MMU67176
Mus musculus tesmin-1 mRNA, complete cds.
U67176
U67176.1 GI:4581558
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 2241)
Sugihara,T., Wadhwa,R., Kaul,S.C. and Mitsui,Y.
A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation
Genomics 57 (1), 130-136 (1999)
99208669
2 (bases 1 to 2241)
Sugihara,T.
Direct Submission
Submitted (19-AUG-1996) NIBH, AIST, Higashi 1-1, Tsukuba, Ibaraki, Japan, 305
Location/Qualifiers

ALIGNMENTS

	1	2241	100.0	2241	10	MMU67176
1	2241	78.0	1861	10	MMU77383	U67176 Mus musculus
2	1747	75.7	2215	10	AB057423	U77383 Mus musculus
3	1695.6	37.5	22856	10	AB057422	AB057423 Mus muscu
4	840.6	28.9	2134	9	HSU86074	UB6074 Homo sapien
5	647.8	7.6	179953	2	AC097959	AC097959 Rattus no
6	169.8	6.5	179953	2	AP003096	AP003096 Homo sapi
7	145.8	5.5	157269	2	AP003096	AP003096 Homo sapi
8	141	4.8	188859	2	AP003732	AP003732 Homo sapi
9	123.6	3.8	214055	2	AC099385	AC099385 Rattus no
10	107.4	3.8	247253	2	AC103272	AC103272 Rattus no
11	85.4	3.8	297288	2	AC106533	AC106533 Rattus no
12	85.4	3.7	187360	9	AC021105	AC021105 Homo sapi
13	83.6	3.4	40893	3	CBRC45E19	AC084631 Caenorhab
14	76.4	3.3	100047	8	AC006081	AC006081 Arabidops
15	74.8	3.3	97714	8	ATF25024	AL078469 Arabidops
16	73.6	3.3	101679	8	ATF19B15	AL078470 Arabidops
17	73.6	3.3	198429	8	ATCHRIV70	AL161574 Arabidops
18	73.6	3.2	94362	8	AC005405	AC005405 BAC F18A1
19	72.8	3.2	170020	8	AP003256	AP003256 Oryza sat
20	72.8	3.2	180206	2	AP003274	AP003274 Oryza sat
21	72.8	3.2	39553	3	CEJCB	282274 Caenorhabdi
22	71.2	3.2	330724	2	CEY67H2	AL022475 Caenorhab
23	71.2	3.1	27555	2	AC012903	AC012903 Drosophil
24	69	3.1	146153	3	AC007452	AC007452 Drosophil
25	69	3.1	159065	3	AC01634	AC01634 Drosophil
26	69	3.1	264646	3	AE003818	AE003818 Drosophil
27	69	3.0	188859	2	AP003732	AP003732 Homo sapi
28	67.6	2.8	128463	2	AP003619	AP003619 Oryza sat
29	63.2	2.8	159947	2	AP003763	AP003763 Oryza sat
30	63.2	2.8	188844	2	AC099384	AC099384 Rattus no
31	63.2	2.7	64971	2	AC107744	AC107744 Mus muscu
32	61.4	2.4	2146	8	AF205142	AF205142 Arabidops
33	53.6	2.4	125020	9	AF429315	AF429315 Homo sapi
34	53.4	2.4	79355	8	AB012247	AB012247 Arabidops
35	53.2	2.4	1793	8	AY046019	AY046019 Arabidops
36	53.2	2.4	2298	8	AF204059	AF204059 Arabidops
37	53	2.4	2431	8	AF206324	AF206324 Arabidops
38	53	2.3	2110	5	AF078111	AF078111 Xenopus 1
39	53	2.3	188844	2	AC099384	AC099384 Rattus no
40	50.8	2.2	158782	2	AC044874	AC044874 Homo sapi
41	50.8	2.2	1072	3	AB024943	AB024943 Tegula pf
42	49.6	2.1	154242	2	AC073373	AC073373 Mus muscu
43	49.2	2.1	163788	2	AC018996	AC018996 Homo sapi
44	47.2	2.1				
45	47	2.1				

```
source 1. .2241
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
651. .1538
CDS
/codon_start=1
/product="tesmin-1"
/protein_id="AAD24566.1"
/db_xref="GI:4581559"
/translation="MVICQLKGQAQMLCINDGARELKHLLPQYDQSPQSELP
KPMWTLVGLRPVPAKLNIITVDNGALPSAVNGAAPPSPALQPPKTLTSGYDCDF
SSGDFNCSNHLRHELFRAIKACLDNPEAFOPKMGKRLGAALKRHSKGCNCK
SGCLKNYCEYEAKIMCSICKIACKNYEESPCKMLMSTPHYNEPDRFSHYLS
PKFSGPPKLRNRQAFSCISWEVVEATCACLQAQEEAEQEHCSPLAEQMLIEFFG
RLSQLHLHLEFSKGLKIE"
BASE COUNT 558 a 554 c 583 g 546 t
ORIGIN

Query Match 100.0%; Score 2241; DB 10; Length 2241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tatctgtggttgcccgccgagcagcagcgtctcagcagcgtcagcacacagagata 60
DB 1 TATCTGTGGGTGGCCCCGGGAGAGGCTCTCAGCAGGCTCAGCACACCAAGGGATA 60

QY 61 cacagtgtgttccctggcctgttgactgtgacttccacccaccctccgcccagcagg 120
DB 61 CACAGTGTGTGTCCTGGCCTGTGGACTGTGACTCCACCCACCTCCGCCCCAGCAGG 120

QY 121 ctaggatagaacccagggccttttgcgtgtctgcagatagctctcagcctggtgatt 180
DB 121 CTAGGATAGAACCAGGGCCTTTTGCCTGTCTGCAGATAGTCTTTCAGGCTGGTAGTT 180

QY 181 ggggttgctggagatttttttctcaaccaagaagccttccattattgagatttt 240
DB 181 GGGGTTGGCTGGGAGATTTTTTTCTTACACCAAGACTTCCATATTAGGAGATTTT 240

QY 241 tcagttgatgatcccccctctgtaagataaaggagcagttctttaaaccctatgtagat 300
DB 241 TCAGTTGATGATCTCCCCCTCTGTAAAGATAAGGGACAGTCTTTAAACCTATGTAGAT 300

QY 301 ttgatgaattctgtctcaaccattatgttaagctatataagcaattctctgaaattgc 360
DB 301 TTTGATGAATTTCTGCTCTCAACCATATTGCTAAAGCTATATAGCAATTCCTTGAATTC 360

QY 361 tataacttaggaacacctctgattctctgcctctacatctcgtagctgaggtatata 420
DB 361 TATATACTTAGAAGAACCTCTGATTCTCTGCTCTACATCTCTAGTGTCTAGGTGTACA 420

QY 421 gggggaaatcatttggtagactccgatgaactactgcaggttccccaaggcagcaagc 480
DB 421 GGGGAAATCATTTTGGTGAGACTCCGATGAACACTACTGCCAGGTTCCCAAGGCAGCAAG 480

QY 481 aagcaagaaaaagcgttgaaatacaagaagcaggtgtgtgtgtgcagcggcagccctg 540
DB 481 AAGCAAGAAAAAGTGTGAAATCAAAAGAACAGAGTGTGTGTGTGCAGGCGGCGAGCCCTG 540

QY 541 aagacagcatttccagccctctgctcaggaatccctgttgcaagttcccatcattcc 600
DB 541 AAGACGACGCTTCCAGGCCCTCTGGCTCAGGAATTCCTGTGTGCAAGTTCCCATCATCC 600

QY 601 agaggcagaggagcctccagctgcccctcggaagaagaagactccagcccatggtgattt 660
DB 601 AGGAGCAGAGGAGGCTCCAGCTGCCCTCGGAAGAAAGACTCCAGCCCCCATGGTGATT 660

QY 661 gtcagctgaagaggcgcagcagatctctgcatagacaactgtgctgagggagctca 720
DB 661 GTCAGCTGAAGAGGCGCCAGATGCTCTGCTAGACAACTGTGCGCGAGGAGGAGCTCA 720

QY 721 aagcgtccatctgctctcagtagcagcagcaggtttccctcagtcagtcagctcc 780
DB 721 AAGCCTCATCTGCTCTCAGTAGCAGCAGGTTTCCCTCAGTCAGTCAGTCAGTCAG 780
```

```
DB 721 AACGGCTCCATCTGCTCTCCTCAGTACGATCACCAGAGCAGTTTCCCTCAGTCAGAGCTCC 780
QY 781 ctaagccaatgacaacttttagtgggaagactcttgcagaccaggaagtaaatctca 840
DB 781 CTAAGCCAATGACAACATTTAGTGGGAAGACTTCTGCCAGTACCAGGGAAGTTAAATCTCA 840
QY 841 tcacacaggttgataatgagctctcccatcagctgtcaatgggggtgccttccctctg 900
DB 841 TCACACAGGTGTGATAATGAGAGCTCTCCCATCAGCTGTCAATGGGGTGCCTTTCCCTCTG 900
QY 901 gacctctctgaagggccacccaaaataactctgtctgggtactgtgactgcttctcca 960
DB 901 GACCTCTCTGCAAGGGCCACCCAAAATAACTCTGTCTGGGTACTGTGACTGCTTCTTCCA 960
QY 961 gggggactcttctcaacagctgcagctgcaacacacctggccatgagctgagcagctca 1020
DB 961 GCGGGGACTTCTGCAACAGCTGCAGCTGCAACAACCTGCGCCATGAGCTGAGCGGCTTCA 1020
QY 1021 aagccataaaggcgtgtcttgatagaaaactcctgaagctttcccaacaaaatgggaaag 1080
DB 1021 AAGCCATAAAGGCGTCTCTGTATAGAAATCCTGAAGCTTTTCCAAACCAAAATGGGAAAG 1080
QY 1081 gccgtctggagctgtctaaacttcgacacagcaaaagggtgcaactgtaagcgtcaggct 1140
DB 1081 GCCGTCTGGAGCTGTAAACTTCGACACAGCAAAAGGGTGCAACTGTAAAGCGCTCAGGCT 1140
QY 1141 gccagaagaactactgtgagtgctatgagggcgaataatcatgtgttcttccattgcaaat 1200
DB 1141 GCCTGAAGAAGTACTCTGAGTGTATGAGGCCAAATCATGTGTTCTTCCATTTGCAAT 1200
QY 1201 gaattcttgcataaactatgaagaagtcacagaacgaataatgctgtagagcacacccc 1260
DB 1201 GCATTCTTGCATAAACTATGAAGAAGTCCAGAACGAAATAATGCTGATGAGCACACCCC 1260
QY 1261 actacatgagcctgggacttttgagcagcagccattattgtccccagcagaagttctcag 1320
DB 1261 ACTACATGAGCCTGGGGACTTTGAGAGCAGCCATTATTGTCCCCAGCCAAAGTTCTCAG 1320
QY 1321 gacctccaaaactgagaaaaaataagcagcagccttctcctgtatctcctgggaagtagtg 1380
DB 1321 GACCTCCAAAACCTGAGAAAAAATAGCAGGSCCTTCTCCTGTATCTCTCGGAAGTAGTGG 1380
QY 1381 aggcacacatgtgcctcctgctgctggccagggttgagaaagcagagcagggagcactgtccc 1440
DB 1381 AGGCCACATGTGCCTGCCTGCTGGCCAGGCTGAGGAAGCAGAGCAGAGCAGTGTTCCTCC 1440
QY 1441 caagctgtgctgagcagatgctcctgagagagtttgagaggtgctgctgcagattctcc 1500
DB 1441 CAAGCTTGGCTGAGCAGATGATCCTGGAGAGGTTTGGAAAGGTGCCCTGTGCAGATTCTCC 1500
QY 1501 acatcgagttcaagttcaagggggtgaaatgtagtagcgtgcaagcgtgtaaggggaa 1560
DB 1501 ACATCGAGTTCAAGTCCAAAGGGGCTGAAATTCAGTAGCTGCAAGCTGGTAAAGGGGAA 1560
QY 1561 tgcctgtgcaagcctcagccctgggaatctgcacagaggaagcgtggtgcccaaggagaga 1620
DB 1561 TGCCTGTGCAAGCCCTCAGCCCTGGGAATCTGCACCGAGGAAGCTGGTGCCTCCAGGAGGA 1620
QY 1621 gcagagggcgcgcacatcattggccaggtcagctgtgaggtctgagtgatctgcatgtaactg 1680
DB 1621 GCAGAGGCCCGGCATCATGCCAGGTCAGCTGTGAGGCTGTGAGTGTGATGTGATGTGACTG 1680
QY 1681 gccagcactactcaagggtatcctaaagtgcagcagcagcagccacctgggagtgagacac 1740
DB 1681 GCCAGCCTACTCAAGGTATCCTTAAAGTGAAGCAAGCAGCAGCAGCCCTGGGATGAGACAC 1740
QY 1741 tggccctctgttccctggggagccctctggggagcctccctgcctgcataaaaaaagggtt 1800
DB 1741 TGGCCCTCTGCTCCCTGGGGAGGCCCTCTGCGGAGCTCCCTGCGCCCTGCATATAAAGAGG 1800
QY 1801 gatttctacttctgttattgtgttgcatttcaaaatgtcttagtagtacctccattcaag 1860
DB 1801 GATTTCTACTTGTGTTATGTGTGCTTTCAATATGCTTAGTAGTACTCCTCATTTCAAG 1860
```

Best Local Similarity 100.0%; Pred. No. 0;				
Matches 1747; Conservative 0; Mismatches 0; Indels 0; Gaps				
Qy	495	gttgaatacaaaagacaggtgtagtggccagggcgagcctgaagacagcagcttcc	554	
Db	115	gTTGAATCAAAAGACAGGTGTAGTGTGCCAGGGCGAGCCCTGAAGACGACGCTTC	174	
Qy	555	caggccccctggctcaggaatacctgtgtcgaagtctccatccccagggcagagag	614	
Db	175	CAGGCCCTCTGGCTCAGGAATCCTGTGCAAGTTCCCATCATCCAGGAGCAGAGAG	234	
Qy	615	gcctcagctgcctcggaagaagactccagccccatggtatattgtcagctgaaaga	674	
Db	235	GCCTCCAGCTGCCTCGGAAGAAGACTCCAGCCCCATGGTATTGTTCAGCTGAAGA	294	
Qy	675	ggcgccagatgctctgtcagtagacaactgtggcgagggagctcaaaagcgtccatctg	734	
Db	295	GGCGCCAGATGCTCTGCATAGACAACTGTGGCGGAGGAGCTCAAGGCTCCATCTG	354	
Qy	735	cttctcagtagcatgataccagagcagtttccctcagtcagagctccctaaagccaatgaca	794	
Db	355	CTTCTCAGTACGATACCAAGACAGTCTTCCCTCAGTCAGAGTCCCCTAAGCAATGACA	414	
Qy	795	acttttagtggaagaactctgcaggtaccagcggaagttaaaatcacaacagattgat	854	
Db	415	ACTTTAGTGGGAAGACTTCTGCCAGTACACAGCGAAGTTAAATCTCATCACAGAGTTGAT	474	
Qy	855	aatggagctctccatcagctgtcaatggggtgcctttccctctggacctgctctgcaa	914	
Db	475	ATYGGAGCTCTCCCATCAGCTGTCAATGGGGTGCCTTTCCCTCTGGAGCTGCTCTGCA	534	
Qy	915	gggcaacccaaaataactctgtctggtagtctgactgtcactgtctccagcgggactctgc	974	
Db	535	GGGCCACCCAAAATACTCTGTCTGGGTACTGTGACTGCTTCTCCAGCGGGACTTCTGC	594	
Qy	975	aaagctgcagctgcacaactgcgccatgagctgcagctgcgtttcaagcaataaagcg	1034	
Db	595	AACAGCTGTCAGCTGCAACCACTTGCGCCATGAGCTCGAGCGCTTCAAGGCCATAAAGCG	654	
Qy	1035	tgtctgtatagaactctgaagctttccaaacaaaatgggaagagcgctctggagct	1094	
Db	655	TGCTCTGATAGAAATCTGAAGCTTTCCAAACCAAAATGGGAAGGCCGCTCTGGAGCT	714	
Qy	1095	gctaaacttcgacacagcaaaaggtgcaactgttaagcgtcaggtgcgtgaagaactac	1154	
Db	715	GCTAAACTTCGACACAGCAAGGTCGAACCTGTAAAGCGCTCAGGCTGCCTGAAGAATAC	774	
Qy	1155	tgtagtgctatagggccaaaaatcatgtgtttctccatttgcaaatgcaattgcttgcaaa	1214	
Db	775	TGTGAGTGTCTATGAGGCCAAAATCATGTGTCTTCTTCCATTTGCAAAATGCAATTCCT	834	
Qy	1215	aactatagaagaagtccagaacgaaaaatgctgatgagcacaccccactacatgagct	1274	
Db	835	AACATATGAAGAAAGTCAGAAACGAAAATGCTGATGAGCACACCCCTACATGAGAGCT	894	
Qy	1275	ggggactttgagagcaccattattgtccccagcaagtctctcaggacctccaaaactg	1334	
Db	895	GGGGACTTTCAGAGCAGCCATTATTGTCCCAGCCCAAGTCTCAGGACCTCCAAAACCTG	954	
Qy	1335	agaaaaataggcagccttctcctgtatctctgggaagttagtggaggccacatgctgcc	1394	
Db	955	AGAAAAAATAGGcAGGCGCTTCTCTGTATCTCTGGGAAGTAGTGGAGGCCACATGTGCC	1014	
Qy	1395	tgcctgctgccccaggttaggaagacagagcaggagcactggttccccagcttgcttag	1454	
Db	1015	TGCCCTGTGCCCCAGGCTGAGGAAGCAGAGCAGGAGCACTGTGTCCCAAGCTTGGCTGAG	1074	
Qy	1455	cagatgatctggagaggtttgaaaggtgcctgtctgcagattctccacatcgagttcaag	1514	
Db	1075	CAGATGATCTCGGAGGAGTTTGGAAAGGTGCCGTCTCGACATTTCTCCACATCAGTTCAAG	1134	
Qy	1515	tccaaggggtgaaaattgagtagcgtgcaagctggttaaggggaatgctctggcaagc	1574	

Db 1006 TGCTTGTATAGAAATCTCTGAAGCTTTTCCAAACCAAAAAATGGGAAAGGCCCTCTCTGGAGCT 1065
 Qy 1095 gctaaactgacacagcaagagtgcaactgtaagcgcagctcagctgcaagaaactac 1154
 Db 1066 GCTAACTTCGACACACAAAGGGTGCACCTGTAGCGCTCAGGCTGCCGTGAAGACTAC 1125
 Qy 1155 tgtgagtgctatgagggccaaatacatgtgttcttccatttgcgaatgattgcttgcga 1214
 Db 1126 TGTGAGTGTATGAGGCCAAATCATGTGTTCTTCCATTGTCAAATGCAATGCTTGTGCAA 1185
 Qy 1215 aactataagaagaagtcagaaacgaaaaatgctgtagagcacacccccactacatgagcct 1274
 Db 1186 AACTATGAAGAAAGTCCAGAACGAAAAATGCTGATGAGCACACCCCTACTACATGAGCCT 1245
 Qy 1275 ggggacttgaagacagccattatttcccccagccagcttctcagacacctccaaaactg 1334
 Db 1246 GGGACATTTGAGACAGCCATTATTGTGCCCCAGCCAAAGTTCTCAGGACCTCCAAAAGTG 1305
 Qy 1335 agaaaaataggcaggccttctctgtatctcctgggaagttagtgaggccacacatgtgcc 1394
 Db 1306 AGAAAAAATAGGCAGGCTTCTCTGTATCTCTGGGAAGTAGTGAGGCCACATGTGCC 1365
 Qy 1395 tgcctgtgccccagggtgaggaagcagagcagagcactgttccccaaagcttgctgag 1454
 Db 1366 TGCTGTGCCCCAGGGTGAGGAAGCAGAGCAGGAGCACGTGTTCCCAAGCTTGGCTGAG 1425
 Qy 1455 cagatgactcgtgagaggttggaggtgctgtcgcagagatttccacatcagagattcaag 1514
 Db 1426 CAGATGATCTTGGAGGAGTTTGGAAAGTGCTGTCGCGAGATTTCCACATCGAGTCAAG 1485
 Qy 1515 tccaggggctgaaaaattgagtagcgtgcaagctggtcaaaagggaatgctgtggaagc 1574
 Db 1486 TCCAAGGGCTGAAAAATGAGTAGCTGCAAGCTGTATAGGGGAATGCCGTGGCAAGC 1545
 Qy 1575 ctacggcctgggaatctgcacccagggaagctggtgccccaggagagcagagccgcga 1634
 Db 1546 CTCAGCCCTGGGAATCTGCACCCAGGAAGCTGTGTGCCAGGAGGAGCAGAGGCCGCCA 1605
 Qy 1635 tcatggcaggtcagctgtgaggtgagtgatctgcatgtagtactgcccagcctactcaa 1694
 Db 1606 TCATGGCCAGGTGACGTGTGAGGTCTCAGTGATCTGATGTTGATGTCGACGCTTACTCAA 1665
 Qy 1695 ggatctcctaagtcaagcagcagcagcagccacctggggatggacactggccctcctgtcc 1754
 Db 1666 GGTATCTAAAGTCAAGCAGCAGCAGCCACCTGGGGATGGACACTGGCCCTCTCTGTC 1725
 Qy 1755 ctggggaggccctctggggactccctgcccctgataaaagaggggtgatttctacttgt 1814
 Db 1726 CTGGGGAGGCCCTCTGGGGACTCCCTGCGCATATAAAAGAGGGGTGATTTTCTACTTGT 1785
 Qy 1815 tgttatgttctgttcaaatgcttagtagtacctccattcagttattatagccag 1874
 Db 1786 TGTATGTGTTGTTTCAAAATGCTTAGTAGTACCTTCCATTCAAGTATTATGAGCCAG 1845
 Qy 1875 cctcaagttagagcagctgcttcttcaagtgagcactgcccacatacacaagtc 1934
 Db 1846 CCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGAGCTGCCCCAATACATACAGTC 1905
 Qy 1935 aggtggccatcaggggtttttccaggccagcctctgacagaggagatagga-ggggggg 1993
 Db 1906 AGGTGGCCATCAGGGG-TCTTCCAGGCCAGGCCCTGTGACAGGGGATATGGGAGGGGGT 1964
 Qy 1994 cgggttagagcgtggttggattttttgg-ttttttcttctgtatttctctt 2052
 Db 1965 CGGGTTAGCTGGGTTGTTGGATTTTGGCGTTTCTCTCTGTTATTTCTGCTT 2024
 Qy 2053 gaagtgaagaaactgtctctgtccaa-ccttttctccataaataactgtgcacggctg 2111
 Db 2025 GAAGTGAGAAAACTGTCTCTGTGCCAACCTCTCTCCATAAATACTGTGTCAGGGTCG 2084
 Qy 2112 cctgctgaccagtcacagtgacccacagaccagagaggtgaggtggtattatgcccac 2171
 Db 2085 CTGCTGACCAGTACAGTGCACCTTCAGACACACAGAAAGTGAGGTGAGGTATTATGCCAC 2144

Qy 2172 acttgtgtttgtgtgagaataaacctttccagactccccaaaaaataaaaaa 2231
 Db 2145 ACTTTGTGTTTGTGTGAGANTAACTTTCCAGACTCCCAAAAAAATAAAAAA 2204
 Qy 2232 aaaaaa 2241
 Db 2205 AAAAAA 2214

RESULT 4
 AB057422 22856 bp DNA linear ROD 26-DEC-2001
 LOCUS Mus musculus gene for tesmin, complete cds, strain:129.
 DEFINITION
 AB057422
 ACCESSION
 AB057422.1 GI:15617419
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (strain:129) DNA.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1 (sites)
 AUTHORS Sutou,S., Miwa,K., Matsuura,T., Kawasaki,Y., Ohinata,Y. and Mitsui,Y.
 TITLE Structure of the tesmin gene encoding a testis-specific persistent protein: a possible multifunctional protein with dynamic changes of localization throughout spermatogenesis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 22856)
 AUTHORS Sutou,S. and Mitsui,Y.
 TITLE Direct Submission
 JOURNAL Submitted (13-MAR-2001) Shizuho Sutou, National Institute of Advanced Industrial Science and Technology (AIST), Institute of Molecular & Cell Biology, Tsukuba Central 6, Higashi 1-1-1, Tsukuba, Ibaraki 305-8566, Japan (E-mail:s-sutou@aist.go.jp, Tel:81-298-61-6052, Fax:81-298-61-9498)
 FEATURES
 Location/Qualifiers
 1..22856
 /organism="Mus musculus"
 /strain="129"
 /db_xref="taxon:10090"
 join(2569..2952,5717..5875,7539..7659,9534..9607, 10577..10668,16100..16190,17665..17802,18550..18725, 20538..20730)
 /gene="tesm"
 join(2569..2952,5717..5875,7539..7659,9534..9607, 10577..10668,16100..16190,17665..17802,18550..18725, 20538..20730)
 /gene="tesm"
 /codon_start=1
 /product="tesmin"
 /protein_id="BAB64934.1"
 /db_xref="GI:15617420"
 /translation="MEDALLGAMTGPDELGAELFSGSERVAFADGLALSPAGGAARDDE LVLADYLGATEPGLPLRALSPPPPAALGVPGLPELRSPPDAAPPAVSF LPSLLPLGAGVPALLPLSAGVRVIVPEIKAGSGVPGSGPDAAPFAQPAQSCCKF PSSQEAEEASCCPRKKDSSPMVICOLKGGQMLCNDNGARELKAHLHLLPOYDDQSF PQSLPKPMITLGRLLPVPKLNLTITQVNDGALPNSAVNGAAPPSPALOGPKITLS GYCDLFSGGDFCNSCCNLRHELERFKAICADLRNPEAFQPMKMGRLGAAKLHRS KGNCKRSGLKNCYCEAKIMCSICCKACKNVEESPERKMLSTPHYMEPGDFE SHYLSPAKFSGPPKLNKQAFSCISWVEVVEATCACLALHGBEAHEHCSPSLAEOM ILBEFGRLSILHIEFKSKGLIE"

BASE COUNT 5724 a 5667 c 5552 g 5913 t
 ORIGIN

Query Match 37.5%; Score 840.6; DB 10; Length 22856;
 Best Local Similarity 99.3%; Pred. No. 3.4e-244;
 Matches 865; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Qy 1344 aggcaggccttctctgtatctctctggaagtagtggaggccacatgtgctgctgctg 1403
 Db 20536 AGGCAGGCTTCTCTGTATCTCTGGAAGTAGTGGAGGCCACATGTGCTGCTGCTG 20595

Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Louisgess, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M.,
 Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogutu, M., Okunuga, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
 Sisson, I., Sodergren, E., Sonaikhe, I., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telirod, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL

2 (bases 1 to 179953)

Worley, K. C.

Direct Submission

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:16901822.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GFSK
 Center clone name: CH230-114G10
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 163172 bases at least Q40
 Consensus quality: 168807 bases at least Q30
 Consensus quality: 174078 bases at least Q20
 Estimated insert size: 160479; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 62 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 6639: contig of 6639 bp in length
 6640 6739: gap of unknown length
 6740 12683: contig of 5944 bp in length
 12684 12783: gap of unknown length
 12784 16867: contig of 4084 bp in length
 16868 16967: gap of unknown length
 16968 22929: contig of 5962 bp in length
 22930 23029: gap of unknown length
 23030 30265: contig of 7236 bp in length
 30266 30365: gap of unknown length
 30366 36389: contig of 6024 bp in length
 36390 36489: gap of unknown length
 36490 42068: contig of 5579 bp in length
 42069 42168: gap of unknown length
 42169 46728: contig of 4560 bp in length
 46729 46828: gap of unknown length
 46829 52208: contig of 5380 bp in length

52209 52308: gap of unknown length
 52309 56609: contig of 4301 bp in length
 56610 56709: gap of unknown length
 56710 59548: contig of 2839 bp in length
 59549 59648: gap of unknown length
 59649 62710: contig of 3062 bp in length
 62711 62810: gap of unknown length
 62811 66227: contig of 3417 bp in length
 66228 66327: gap of unknown length
 66328 69089: contig of 2762 bp in length
 69090 69189: gap of unknown length
 69190 72645: contig of 3456 bp in length
 72646 72745: gap of unknown length
 72746 76817: contig of 4072 bp in length
 76818 76917: gap of unknown length
 80040 80040: contig of 3123 bp in length
 80041 80140: gap of unknown length
 80141 83579: contig of 3439 bp in length
 83580 83679: gap of unknown length
 83680 86465: contig of 2786 bp in length
 86466 86565: gap of unknown length
 86566 90515: contig of 3950 bp in length
 90516 90615: gap of unknown length
 90616 94222: contig of 3607 bp in length
 94223 94322: gap of unknown length
 94323 97924: contig of 3602 bp in length
 97925 98024: gap of unknown length
 98025 98751: contig of 1727 bp in length
 98752 99851: gap of unknown length
 99852 102642: contig of 2791 bp in length
 102643 102742: gap of unknown length
 102743 106798: contig of 4056 bp in length
 106799 106898: gap of unknown length
 106899 110343: contig of 3445 bp in length
 110344 110443: gap of unknown length
 110444 112934: contig of 2491 bp in length
 112935 113034: gap of unknown length
 113035 116034: contig of 3000 bp in length
 116035 116134: gap of unknown length
 116135 119554: contig of 3420 bp in length
 119555 119655: gap of unknown length
 119656 122029: contig of 2375 bp in length
 122030 122129: gap of unknown length
 122130 125047: contig of 2917 bp in length
 125047 125146: gap of unknown length
 125147 127444: contig of 2298 bp in length
 127445 127544: gap of unknown length
 127545 129251: contig of 1707 bp in length
 129252 129351: gap of unknown length
 129352 131551: contig of 2200 bp in length
 131552 131651: gap of unknown length
 131652 133716: contig of 2065 bp in length
 133717 133816: gap of unknown length
 133817 136695: contig of 2879 bp in length
 136696 136795: gap of unknown length
 136796 138444: contig of 1649 bp in length
 138445 138544: gap of unknown length
 138545 140364: contig of 1820 bp in length
 140365 140464: gap of unknown length
 140465 142056: contig of 1592 bp in length
 142057 142156: gap of unknown length
 142157 144300: contig of 2144 bp in length
 144301 144400: gap of unknown length
 144401 145630: contig of 1230 bp in length
 145631 145730: gap of unknown length
 145731 147565: contig of 1835 bp in length
 147566 147665: gap of unknown length
 147666 149525: contig of 1860 bp in length
 149526 149625: gap of unknown length
 149626 151295: contig of 1670 bp in length
 151296 151395: gap of unknown length
 151396 152702: contig of 1307 bp in length
 152703 152802: gap of unknown length

```

* 152803 154292: contig of 1490 bp in length
* 154293 154392: gap of unknown length
* 154393 155990: contig of 1598 bp in length
* 155991 156090: gap of unknown length
* 156091 157685: contig of 1995 bp in length
* 157686 157785: gap of unknown length
* 157786 159784: contig of 1999 bp in length
* 159785 159884: gap of unknown length
* 159885 161347: contig of 1463 bp in length
* 161348 161447: gap of unknown length
* 161448 162678: contig of 1231 bp in length
* 162679 162778: gap of unknown length
* 162779 163996: contig of 1218 bp in length
* 163997 164096: gap of unknown length
* 164097 166978: contig of 2882 bp in length
* 166979 167078: gap of unknown length
* 167079 168859: contig of 1781 bp in length
* 168860 168959: gap of unknown length
* 168960 170817: contig of 1858 bp in length
* 170818 170917: gap of unknown length
* 170918 172018: contig of 1101 bp in length
* 172019 172118: gap of unknown length

Query Match      6.5%; Score 145.8; DB 2; Length 179953;
Best Local Similarity 92.7%; Pred. No. 4.8e-32;
Matches 153; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 495 gttgaatacaagacagtgtagtgcagcgccgagcctgaagcagcgttcc 554
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 115559 GTAGAATCAAGAGCAGTGCGGTGTGCCAGGAGCAGGCTCAAGATCAGTTC 115500

QY 555 cagggccctgtgctcaggaatcctgttgcagctccatccagcagcagggag 614
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 115499 CAGGCCCTCTGCTCAGGAATCCTGCTCAAGTTCTATCCAGGAGGAGGAG 115440

QY 615 gcttcagctgcctcgggaagaagactccagcccatcgtgatt 659
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 115439 GCCTCGGCTGCGCCCGGAAAAAGACTCCAGCCCATGTAAC 115395

RESULT 8
AP003096/c      157269 bp      DNA      linear      HTG 18-JAN-2001
LOCUS      Homo sapiens chromosome 11 clone CTD-2007L18 map 11q, WORKING DRAFT
DEFINITION      SEQUENCE, 32 unordered pieces.
ACCESSION      AP003096.1 GI:1275357
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      Homo sapiens DNA, clone:CTD-2007L18.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157269)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 157,269 genomic DNA of 11q
Published Only in DataBase (2001) In press
2 (bases 1 to 157269)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (16-JAN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
TEL:81-45-503-9111, Fax:81-45-503-9170)
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Humdraft11

```

Center clone name: CTD-2007L18

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator Er-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 147566 bases at least Q40

Consensus quality: 151779 bases at least Q30

Consensus quality: 153479 bases at least Q20

Insert size: 154169; sum-of-contigs

Quality coverage: 8.32x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 11499 contig of 11499 bp in length
11600 19042 contig of 7443 bp in length
19143 26427 contig of 7285 bp in length
26528 38740 contig of 12213 bp in length
38841 55483 contig of 16643 bp in length
55584 67548 contig of 11965 bp in length
67649 74728 contig of 7080 bp in length
74829 79575 contig of 4747 bp in length
79676 88957 contig of 9282 bp in length
89058 93854 contig of 4797 bp in length
93955 98914 contig of 4960 bp in length
99015 102673 contig of 3659 bp in length
102714 105297 contig of 2524 bp in length
105398 110377 contig of 4980 bp in length
110478 117540 contig of 7063 bp in length
117641 120176 contig of 2536 bp in length
120277 125707 contig of 5431 bp in length
125808 128247 contig of 2440 bp in length
128348 131115 contig of 2768 bp in length
131216 134071 contig of 2856 bp in length
134172 136288 contig of 2117 bp in length
136389 139198 contig of 2810 bp in length
139299 141579 contig of 2281 bp in length
141680 143825 contig of 2146 bp in length
143926 146191 contig of 2266 bp in length
146292 148382 contig of 2091 bp in length
148483 149666 contig of 1184 bp in length
149767 151747 contig of 1981 bp in length
151848 153751 contig of 1904 bp in length
153852 154877 contig of 1026 bp in length
154978 156158 contig of 1181 bp in length
156259 157269 contig of 1011 bp in length.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 11499: contig of 11499 bp in length
11500 11599: gap of 100 bp
11600 19042: contig of 7443 bp in length
19043 19142: gap of 100 bp
19143 26427: contig of 7285 bp in length
26428 26527: gap of 100 bp
26528 38740: contig of 12213 bp in length
38741 38840: gap of 100 bp
38841 55483: contig of 16643 bp in length
55484 55583: gap of 100 bp
55584 67548: contig of 11965 bp in length
67549 67648: gap of 100 bp
67649 74728: contig of 7080 bp in length
74729 74828: gap of 100 bp
74829 79575: contig of 4747 bp in length

```

[illegible]

```

LOCUS      AP003096                      137269 bp      DNA
DEFINITION Homo sapiens chromosome 11 clone CTD-2007L18 map llg, WORKING DRAFT
            SEQUENCE, 32 unordered pieces.
ACCESSION  AP003096.1 GI:12275357
VERSION    HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS   AP003096
SOURCE     Homo sapiens
            Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 157269)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE     Homo sapiens 157,269 genomic DNA of llg
JOURNAL    Published Only in DataBase (2001) In press
REFERENCE  2 (bases 1 to 157269)
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
AUTHORS   Direct Submission
TITLE     Submitted (16-JAN-2001) Masahira Hattori, The Institute of Physical
JOURNAL    and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            ----- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
COMMENT

```



```

/misc_feature /note="assembly_fragment"
74829. .79575
/misc_feature /note="assembly_fragment"
79676. .88957
/misc_feature /note="assembly_fragment"
89058. .93854
/misc_feature /note="assembly_fragment"
93955. .98914
/misc_feature /note="assembly_fragment"
99015. .102673
/misc_feature /note="assembly_fragment"
102774. .105297
/misc_feature /note="assembly_fragment"
105398. .110377
/misc_feature /note="assembly_fragment"
110478. .117540
/misc_feature /note="assembly_fragment"
117641. .120176
/misc_feature /note="assembly_fragment"
120277. .125707
/misc_feature /note="assembly_fragment"
125808. .128247
/misc_feature /note="assembly_fragment"
128348. .131115
/misc_feature /note="assembly_fragment"
131216. .134071
/misc_feature /note="assembly_fragment"
134172. .136288

```

```

Query Match 5.5%; Score 123.6; DB 2; Length 157269;
Best Local Similarity 72.9%; Pred. No. 2.7e-25;
Matches 159; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1164 tatgaggccaaatcatgttttccattgcaaatcattgttgcataaaactatgaa 1223
Db 3656 TACCAGGCCCAATATATGTCTTCTATTGCAAAATGATGGTGCATAAATATGAA 36625

QY 1224 gaaatccagaaacgaaatgctgtagcacacccactacatgagcctggggacttt 1283
Db 36626 GAAAGCCCAAGCAAGACATATGATGATGCAAACTACATGACAGTGGAGTTG 36685

QY 1284 gagagcagcattatttctccagccagcagttctcagagctccaaactgagaaaaat 1343
Db 36686 GAAAGCCCAAGCAAGACATATGATGATGCAAACTACATGACAGTGGAGTTG 36745

QY 1344 aggcagcctctctgtatctctcctggagtagtga 1381
Db 36746 AGGTAGTACTTTGAACCTACCTCTGCTAATAGAGGA 36783

```

```

RESULT 10
LOCUS AP003732 188859 bp DNA linear HTG 12-JUN-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-682D22 map 11q, WORKING DRAFT
SEQUENCE, 39 unordered pieces.
ACCESSION AP003732
VERSION AP003732.1 GI:14349298
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-682D22.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 188859)
Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
Fujiyama.A., Yada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
Direct Submission
Submitted (11-JUN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)

```

```

Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: Humdraft11
Center clone name: RP11-682D22
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 181083 bases at least Q40
Consensus quality: 184471 bases at least Q30
Consensus quality: 184447 bases at least Q20
Insert size: 185059; sum-of-contigs
Quality coverage: 7.9% in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 39 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 19831 contig of 19831 bp in length
19932 31497 contig of 11566 bp in length
31598 44647 contig of 13050 bp in length
44748 54198 contig of 9451 bp in length
54299 63363 contig of 9065 bp in length
63464 71678 contig of 8215 bp in length
71779 79960 contig of 8182 bp in length
80061 87061 contig of 7001 bp in length
87162 93934 contig of 6773 bp in length
94035 101099 contig of 7065 bp in length
101200 108623 contig of 7424 bp in length
108724 114505 contig of 5782 bp in length
114606 119720 contig of 5115 bp in length
119821 124688 contig of 4868 bp in length
124789 130420 contig of 5632 bp in length
130521 134923 contig of 4403 bp in length
135024 138158 contig of 3135 bp in length
138259 143086 contig of 4828 bp in length
143187 146616 contig of 3430 bp in length
146717 150135 contig of 3419 bp in length
150236 153465 contig of 3230 bp in length
153566 156263 contig of 2698 bp in length
156364 158882 contig of 2519 bp in length
158983 161109 contig of 2127 bp in length
161210 164035 contig of 2826 bp in length
164136 166612 contig of 2477 bp in length
166713 168665 contig of 1953 bp in length
168766 171614 contig of 2849 bp in length
171715 173550 contig of 1836 bp in length
173651 174290 contig of 640 bp in length
174391 176225 contig of 1835 bp in length
176326 178179 contig of 1854 bp in length
178280 179640 contig of 1361 bp in length
179741 181689 contig of 1949 bp in length
181790 183060 contig of 1271 bp in length
183161 184208 contig of 1048 bp in length
184309 185664 contig of 1356 bp in length
185765 187616 contig of 1852 bp in length
187717 188859 contig of 1143 bp in length.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 39 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 19831: contig of 19831 bp in length

* 19932 31497: gap of 100 bp

* 19932 31497: contig of 11566 bp in length

*	31498	31597:	gap of	100 bp
*	31598	44647:	contig of 13050 bp	in length
*	44648	44747:	gap of	100 bp
*	44748	54198:	contig of 9451 bp	in length
*	54199	54298:	gap of	100 bp
*	54299	63363:	contig of 9065 bp	in length
*	63364	63463:	gap of	100 bp
*	63464	71678:	contig of 8215 bp	in length
*	71679	71778:	gap of	100 bp
*	71779	79960:	contig of 8182 bp	in length
*	79961	80060:	gap of	100 bp
*	80061	87061:	contig of 7001 bp	in length
*	87062	87161:	gap of	100 bp
*	87162	93934:	contig of 6773 bp	in length
*	93935	94034:	gap of	100 bp
*	94035	101099:	contig of 7065 bp	in length
*	101100	101199:	gap of	100 bp
*	101200	108623:	contig of 7424 bp	in length
*	108624	108723:	gap of	100 bp
*	108724	114505:	contig of 5782 bp	in length
*	114506	114605:	gap of	100 bp
*	114606	119720:	contig of 5115 bp	in length
*	119721	119820:	gap of	100 bp
*	119821	124688:	contig of 4868 bp	in length
*	124689	124788:	gap of	100 bp
*	124789	130420:	contig of 5632 bp	in length
*	130421	130520:	gap of	100 bp
*	130521	134923:	contig of 4403 bp	in length
*	134924	135023:	gap of	100 bp
*	135024	138158:	contig of 3135 bp	in length
*	138159	138258:	gap of	100 bp
*	138259	143086:	contig of 4828 bp	in length
*	143087	143186:	gap of	100 bp
*	143187	146616:	contig of 3430 bp	in length
*	146617	146716:	gap of	100 bp
*	146717	150135:	contig of 3419 bp	in length
*	150136	150235:	gap of	100 bp
*	150236	153465:	contig of 3230 bp	in length
*	153466	153565:	gap of	100 bp
*	153566	156263:	contig of 2698 bp	in length
*	156264	156363:	gap of	100 bp
*	156364	158882:	contig of 2519 bp	in length
*	158883	158982:	gap of	100 bp
*	158983	161109:	contig of 2127 bp	in length
*	161110	161209:	gap of	100 bp
*	161210	164035:	contig of 2826 bp	in length
*	164036	164135:	gap of	100 bp
*	164136	166612:	contig of 2477 bp	in length
*	166613	166712:	gap of	100 bp
*	166713	168665:	contig of 1953 bp	in length
*	168666	168765:	gap of	100 bp
*	168766	171614:	contig of 2849 bp	in length
*	171615	171714:	gap of	100 bp
*	171715	173550:	contig of 1836 bp	in length
*	173551	173650:	gap of	100 bp
*	173651	174290:	contig of 640 bp	in length
*	174291	174390:	gap of	100 bp
*	174391	176225:	contig of 1835 bp	in length
*	176226	176325:	gap of	100 bp
*	176326	178179:	contig of 1854 bp	in length
*	178180	178279:	gap of	100 bp
*	178280	179640:	contig of 1361 bp	in length
*	179641	179740:	gap of	100 bp
*	179741	181689:	contig of 1949 bp	in length
*	181690	181789:	gap of	100 bp
*	181790	183060:	contig of 1271 bp	in length
*	183061	183160:	gap of	100 bp
*	183161	184208:	contig of 1048 bp	in length
*	184209	184308:	gap of	100 bp
*	184309	185664:	contig of 1356 bp	in length
*	185665	185764:	gap of	100 bp
*	185765	187616:	contig of 1852 bp	in length
*	187617	187716:	gap of	100 bp

FEATURES	* 187717	188859: contig of 1143 bp in length.	
source	Location/Qualifiers		
	1. 188859		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="11"		
	/map="11q"		
misc_feature	/clone="RP11-682D22"		
	1. 19831		
	/note="assembly_fragment"		
misc_feature	19932..31497		
	/note="assembly_fragment clone_end:SP6 vector_side:left"		
misc_feature	31598..44647		
	/note="assembly_fragment"		
misc_feature	44748..54198		
	/note="assembly_fragment"		
misc_feature	54299..63363		
	/note="assembly_fragment"		
misc_feature	63464..71678		
	/note="assembly_fragment"		
misc_feature	71779..79960		
	/note="assembly_fragment"		
misc_feature	80061..87061		
	/note="assembly_fragment"		
misc_feature	87162..93934		
	/note="assembly_fragment"		
misc_feature	94035..101099		
	/note="assembly_fragment"		
misc_feature	101200..108623		
	/note="assembly_fragment"		
misc_feature	108724..114505		
	/note="assembly_fragment"		
misc_feature	114606..119720		
Query Match	4.88;	Score 107.4; DB 2; Length 188859;	
Best Local Similarity	78.2%;	Pred. No. 2.5e-20;	
Matches 129;	Conservative 0;	Mismatches 36; Indels 0; Gaps 0;	
Qy 495	gttgaatcaaaagaagcagggtggtagtgccaggcggcagccctgaagcagcagcttc 554		
Db 141024	GTTGAATCAGGAAGCAGGTGGTACTACTACAGTAAATATCGGGAAGAGCAACTTTG 141083		
Qy 555	caggccctctggctcagggaatcctgttgcgaagttcccatcatccccaggaggcagaggag 614		
Db 141084	CAGAATCTCTTGCTCAGGAATCTGTTGCAAGTTCCATCGTCCAGGAACATAGAGGAT 141143		
Qy 615	gcctccagctgcctcctcggagaagaagactccagcccatgtgtgatt 659		
Db 141144	GCCTCCCTGCTTCTTCTTAAGAAAGATTTCCAAACCAATGTTAAATT 141188		

RESULT	11
AC099385/c	
LOCUS	
DEFINITION	AC099385 214055 bp DNA linear HTG 21-DEC-2001 Rattus norvegicus chromosome Rf4 clone CH230-182G12, *** SEQUENCING IN PROGRESS ***, 90 unordered pieces.
ACCSSION	AC099385
VERSION	AC099385.2 GI:17974832
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 214055) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
AUTHORS	

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Dunn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
 Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, J., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Loulseghe, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogih, M., Okuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
 Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmari, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczkyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 214055)
 Worley, K.C.
 Direct Submission
 Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 21, 2001 this sequence version replaced gi:16901968.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GKDB
 Center clone name: CH230-182G12
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 148220 bases at least Q40
 Consensus quality: 157150 bases at least Q30
 Consensus quality: 162432 bases at least Q20
 Estimated insert size: 127504; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 90 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 9985: contig of 9985 bp in length
 * 9986 10085: gap of unknown length
 * 10086 15886: contig of 5801 bp in length
 * 15887 15986: gap of unknown length
 * 15987 22802: contig of 6816 bp in length

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 187360)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 187360)
Waterston,R.H.
Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 187360)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 3, 2001 this sequence version replaced gi:13431110.

REFERENCE
AUTHORS
TITLE
JOURNAL

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_NH0163017

REFERENCE
AUTHORS
TITLE
JOURNAL

Center project name: H_NH0163017

COMMENT

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBac3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-163017; actual end is at base position 187360 of RP11-163017.

The region between bases 32935-33083 is single stranded and the sequence fidelity can not be guaranteed. RP11-163017 contains an unresolved dinucleotide repeat between bases 47950-48140 where the sequence fidelity can not be guaranteed. Digest information suggests that approximately 88 bases are missing from the current assembly.

FEATURES

source

Location/Qualifiers
1..187360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-163017"

```

/clone_lib="RPCI-11"
1..85 misc_feature
/notes="similar to EST BE614739 (NID:g9896336)"
1..85 misc_feature
/notes="similar to EST BE786771 (NID:g10207969)"
1..85 misc_feature
/notes="similar to EST AV610323 (NID:g9745993)"
1..85 misc_feature
/notes="similar to EST AU120774 (NID:g10936009)"
1..85 misc_feature
/notes="similar to EST A1178109 (NID:g3728747)"
1..85 misc_feature
/notes="similar to EST BE074520 (NID:g8422511)"
1..85 misc_feature
/notes="similar to EST AA397735 (NID:g2050626) zt86a05.r1"
209..800 misc_feature
/notes="similar to EST BG432741 (NID:g13339247)"
227..725 misc_feature
/notes="similar to EST BF911522 (NID:g12302980)"
649..800 misc_feature
/notes="similar to EST A1178109 (NID:g3728747)"
649..778 misc_feature
/notes="similar to EST BE786771 (NID:g10207969)"
649..800 misc_feature
/notes="similar to EST AU120774 (NID:g10936009)"
649..800 misc_feature
/notes="similar to EST AV610323 (NID:g9745993)"
649..800 misc_feature
/notes="similar to EST AA397735 (NID:g2050626) zt86a05.r1"
649..800 misc_feature
/notes="similar to EST BE614739 (NID:g9896336)"
649..777 misc_feature
/notes="similar to EST BE074520 (NID:g8422511)"
651..800 misc_feature
/notes="similar to EST BG703239 (NID:g13975373)"
655..800 misc_feature
/notes="similar to EST T32984 (NID:g615082)"
1023..1310 repeat_region
/rpt_family="Alu"
1309..1334 repeat_region
/rpt_family="(TTAA)n"
1678..1878 repeat_region
/rpt_family="MIR"
1938..1965 repeat_region
/rpt_family="AT-rich"
2379..2416 repeat_region
/rpt_family="AT-rich"
2399..2711 repeat_region
/rpt_family="Alu"
3034..3091 repeat_region
/rpt_family="MIR"
3295..3779 repeat_region
/rpt_family="L1"
3784..3834 repeat_region
/rpt_family="L1"
3913..4049 repeat_region
/rpt_family="MIR"
4261..4318 misc_feature
/notes="similar to EST AA397735 (NID:g2050626) zt86a05.r1"
4261..4340 misc_feature
/notes="similar to EST BG432741 (NID:g13339247)"
4261..4340 misc_feature
/notes="similar to EST AU120774 (NID:g10936009)"
4261..4340 misc_feature
/notes="similar to EST BE614739 (NID:g9896336)"
4261..4340 misc_feature
/notes="similar to EST A1178109 (NID:g3728747)"
4261..4340 misc_feature
/notes="similar to EST BG703239 (NID:g13975373)"
4261..4340 misc_feature
/notes="similar to EST T32984 (NID:g615082)"
4261..4340 misc_feature
/notes="similar to EST AV610323 (NID:g9745993)"

```

COMMENT
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by: Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspieth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

FEATURES	Location/Qualifiers
SOURCE	1. 40893
	/organism="Caenorhabditis b
	/strain="Gujarat G16"
	/db.xref="taxon:6238"
	/clone="G45E19"
BASE COUNT	13086 a 7791 c 7927 g 12089 t
ORIGIN	

Query Match	3.4%	Score	76.4	DB	3	Length	40893
Best Local Similarity	58.4%	Prod. NO.	4.8e-11				
Matches	181	Conservative	0	Mismatches	111	Indels	18
Gaps	2						
Qy	932	tctgtctgggtactgtgactcttcacagcggggaactctgcaacagctgcagctg---	988				
Db	37990	TCTGAACATCTACTGTGACTGCTTCGCAACGAGGAGTTTTGCCGTGACTGCAACTGC	38049				
Qy	989	-----caacaacctgcgcctagctcgagcgcttcaagcgcataaagcgctgtct	1039				
Db	38050	GGACTGCCACAACAATATAGATTATGACAGCCACGATCAAAAGCCATTCGTCAGCTTT	38109				
Qy	1040	tgatagaatcctgaagctttcccaaccaaaaatgggaaagccctcgtggagctgctaa	1099				
Db	38110	GGAGCGAAATCCAAATGCTTTCAAGCCGAAATCGAATAGCTCGTGGTGGAAACCGCG	38169				
Qy	1100	acttcgacacagc-----aaagggtgcaactgtaagcgctcaggctgcctgaaagaacta	1153				
Db	38170	TATTGAACGCTTGCAATCAAAAGGGATGCCACTGCAAGAAAAACGGTTCCTGAAAAACTA	38229				
Qy	1154	ctgtgagtgctatgagcccaaaatcatgttcttcatttgcaaatgcattgcttgc	1213				
Db	38230	TGGCAGTGCTACGAAAGCAAAAGTGCCCTGTACTGATCGATGCAAGTCAAAAGCGTGCA	38289				
Qy	1214	aaactatgaa	1223				
Db	38290	GAATACGGAA	38299				

Search completed: July 11, 2002, 11:10:47
Job time: 10564 sec

[illegible]

RESULT 15
 CBRG45E19
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Caenorhabditis briggsae.
 Caenorhabditis briggsae
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 40893)
 The C. briggsae Genome Sequencing Center.
 The C. briggsae Genome Sequencing Project
 unpublished
 JOURNAL
 REFERENCE
 2 (bases 1 to 40893)
 AUTHORS
 TITLE
 Direct Submission
 Submitted (04-NOV-2000)
 Department of Genetics, Washington

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:14:43 ; Search time 479.19 seconds
(without alignments)
8029.394 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 2241

Sequence: 1 tatctctgtgtgtgccccg.....aaaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2241	100.0	2241	21	AAZ88155
2	1747	78.0	1861	21	AAZ88157
3	647.8	28.9	2134	21	AAZ88156
4	248.4	11.1	445	22	AAZ81300
5	168	7.5	3495	22	AAK94268
6	157.4	7.0	3533	20	AAV72867
7	120.4	5.4	250	22	AAH35563
8	105.4	4.7	1093	23	AAZ82145
9	92.8	4.1	3164	23	ABL10437
					Mouse testis speci
					Mouse testis speci
					Human testis speci
					Human reproductive
					Human full-length
					Human lin-54 homol
					Human colon cancer
					DNA encoding novel
					Drosophila melanog

10	71.2	3.2	1503	20	AAV72865	Caenorhabditis ele
11	69	3.1	8372	23	ABL10436	Drosophila melanog
12	46.4	2.1	732	23	ABL24143	Drosophila melanog
13	45.6	2.0	372	21	AAZ88155	Human secreted pro
14	45.2	2.0	387	22	AAI89088	Human polynucleoti
15	44.8	2.0	1696	21	AAF16214	Human prostate can
16	44.8	2.0	3891	21	AAZ76424	Human ORFX ORF1979
17	44.2	2.0	615	21	AAZ80287	Human colon cancer
18	44.2	2.0	819	19	AAV63198	cDNA from clone ga
19	44	2.0	2132	20	AAZ33816	Coding sequence fo
20	44	2.0	2794	23	ABL24142	Drosophila melanog
21	44	2.0	5529	23	ABL24146	Drosophila melanog
22	43.8	2.0	768	22	AAZ80893	Human breast cance
23	43.6	1.9	964	15	AAQ68357	Apo-lipoprotein AI
24	43.4	1.9	397	22	AAI87874	Human polynucleoti
25	43.4	1.9	2043	21	AAZ22359	Human secreted pro
26	43.4	1.9	2043	22	AAH33065	Human colon cancer
27	43	1.9	243	22	AAZ80490	Human breast cance
28	43	1.9	658	21	AAZ16322	Human colon cancer
29	43	1.9	1241	21	AAZ17553	Human breast and o
30	42.8	1.9	3024	11	AAQ04049	Sequence of clone
31	42.8	1.9	3024	15	AAQ57397	Human ICAM-1. Hom
32	42.8	1.9	3024	17	AAZ11672	Intracellular adhe
33	42.8	1.9	3024	18	AAZ62835	Human ICAM-1 cDNA
34	42.8	1.9	3024	18	AAZ47957	Human ICAM-1 (HL-6
35	42.8	1.9	3024	19	AAV70950	Intracellular adhe
36	42.8	1.9	3113	10	AAZ80995	HL-cDNA clone enco
37	42.6	1.9	1162	22	AAH33698	Human colon cancer
38	42.6	1.9	4123	22	AAZ72784	Human immune/haema
39	42.4	1.9	738	21	AAZ16171	Human prostate can
40	42.4	1.9	1264	21	AAZ77827	Human cancer assoc
41	42.4	1.9	6668	22	AAZ46418	Tumour suppressor
42	42.4	1.9	6668	24	ABL33219	Human immune syste
43	42.2	1.9	430	22	AAI84605	Human polynucleoti
44	42.2	1.9	772	22	AAI96084	Human neuroblastom
45	42.2	1.9	818	22	AAI94242	Human neuroblastom

ALIGNMENTS

RESULT 1

AAZ88155

ID AAZ88155 standard; cDNA; 2241 BP.

XX AC AAZ88155;

XX DT 25-APR-2000 (first entry)

XX DE Mouse testis specific factor tesmin encoding cDNA SEQ ID NO:1.

XX KW Testis specific factor; tesmin; cell death; regulation; spermatocyte; differentiation regulatory factor; male germ cell regulatory actor;

XX KW germ cell differentiation; sterility; ss.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

XX FT CDS 651..1538

XX FT /*tag= a

XX FT /product= "tesmin"

XX PN WO200004147-A1.

XX PD 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-JP03859.

XX PR 17-JUL-1998; 98JP-0219856.

XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX WPI: 2000-147785/13.
DR P-PSDB; AAY6463.

XX New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility -
PT

XX Claim 3; Page 33-37; 63pp; Japanese.

PS The present sequence encodes a male germ cell regulatory factor
CC expressed specifically in spermatocytes, designated tesmin.
CC can be used in the investigation of the mechanisms of germ cell
CC differentiation and sterility.

XX Sequence 2241 BP; 558 A; 554 C; 583 G; 546 T; 0 other;

Query Match 100.0%; Score 2241; DB 21; Length 2241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	tatctctggtggtgccccggcagcagcagcgtctcagcagcgtcagcaccacacaggata	60
DB	1	tatctctggtggtgccccggcagcagcagcgtctcagcagcgtcagcaccacacaggata	60
QY	61	cacagtgtgtctcctggcctgtgactgtgacttccacccacaccccccagcagg	120
DB	61	cacagtgtgtctcctggcctgtgactgtgacttccacccacaccccccagcagg	120
QY	121	ctaggatgagaaacccggccttttgcgtgtctgcagatgacttccagcgtgtagttt	180
DB	121	ctaggatgagaaacccggccttttgcgtgtctgcagatgacttccagcgtgtagttt	180
QY	181	ggggtggtcggagatgttttttttccacacaaagacttccattattgagatttt	240
DB	181	ggggtggtcggagatgttttttttccacacaaagacttccattattgagatttt	240
QY	241	tcagtgtgatctccccctctgtaagataaggagcagcttcttaaacctatgtagtt	300
DB	241	tcagtgtgatctccccctctgtaagataaggagcagcttcttaaacctatgtagtt	300
QY	301	tttgatgaaattctgcttccacacattgttgcgtgtctgcagatgacttccatgaaattgc	360
DB	301	tttgatgaaattctgcttccacacattgttgcgtgtctgcagatgacttccatgaaattgc	360
QY	361	tataacttagagaaacctctgacttctcctcctctacacctgactgactggtgtaca	420
DB	361	tataacttagagaaacctctgacttctcctcctctacacctgactgactggtgtaca	420
QY	421	gggggaaatcattttggtagactccgatgaactactgcccaggttcccaggcagcaagc	480
DB	421	gggggaaatcattttggtagactccgatgaactactgcccaggttcccaggcagcaagc	480
QY	481	aagcaaaaaaagtgttgaatacaaaagacaggtgtgtagtgcagggcgagccctg	540
DB	481	aagcaaaaaaagtgttgaatacaaaagacaggtgtgtagtgcagggcgagccctg	540
QY	541	aagacgagctttccagggccctctgctcaggaatcctgttgcaggttcccatcctcc	600
DB	541	aagacgagctttccagggccctctgctcaggaatcctgttgcaggttcccatcctcc	600
QY	601	aggaggcagaggagcctccagctgcccctggaaagaagactccagcccctatggtatt	660
DB	601	aggaggcagaggagcctccagctgcccctggaaagaagactccagcccctatggtatt	660
QY	661	gtcagctgaaaggaggcgcccagatgctctgcatagacaaactgtgagcagaggagctca	720
DB	661	gtcagctgaaaggaggcgcccagatgctctgcatagacaaactgtgagcagaggagctca	720
QY	721	aagcgtccatctgcttccctcagtagcatgacagagcaggttccctcagtcagagctcc	780
DB	721	aagcgtccatctgcttccctcagtagcatgacagagcaggttccctcagtcagagctcc	780

QY	781	ctaagccaatgacaacttttagtgggaagacttctgcccagtagcagcgaagttaaattctca	840
DB	781	ctaagccaatgacaacttttagtgggaagacttctgcccagtagcagcgaagttaaattctca	840
QY	841	tcacacagttgataatgagcgtctcccatcagctgctcaatggggtccttccctctg	900
DB	841	tcacacagttgataatgagcgtctcccatcagctgctcaatggggtccttccctctg	900
QY	901	gacctgctgcaaggccaccacaaataactctgtctgggtactgtgactgotttctca	960
DB	901	gacctgctgcaaggccaccacaaataactctgtctgggtactgtgactgotttctca	960
QY	961	gcggggaacttctgcaacagctgcagctgcacaaacctgcgcacatgcagcgtctca	1020
DB	961	gcggggaacttctgcaacagctgcagctgcacaaacctgcgcacatgcagcgtctca	1020
QY	1021	aagccataaaggcgtgtctttagatagaatccctgaagctttccacacaaatggggaag	1080
DB	1021	aagccataaaggcgtgtctttagatagaatccctgaagctttccacacaaatggggaag	1080
QY	1081	gcgctggtggcgtcgttaacttcgacacagcaaaagggtgcaactgtaagcgtcaggct	1140
DB	1081	gcgctggtggcgtcgttaacttcgacacagcaaaagggtgcaactgtaagcgtcaggct	1140
QY	1141	gocctgaagaactactgtgagtgctatgagggcacaatcatgtgttctccattgcaaat	1200
DB	1141	gocctgaagaactactgtgagtgctatgagggcacaatcatgtgttctccattgcaaat	1200
QY	1201	gcattgttgcacaaactatgaagaagtcacagacgaaatgctgagtagcagacacccc	1260
DB	1201	gcattgttgcacaaactatgaagaagtcacagacgaaatgctgagtagcagacacccc	1260
QY	1261	actacatgagcctgggactttgagcagcagccattatttgcctccagcagcagcttccag	1320
DB	1261	actacatgagcctgggactttgagcagcagccattatttgcctccagcagcagcttccag	1320
QY	1321	gacctccaaactgagaaaaatagcagcagccttctcctgtatctcctgggaagtgtg	1380
DB	1321	gacctccaaactgagaaaaatagcagcagccttctcctgtatctcctgggaagtgtg	1380
QY	1381	agccacatgtgctgctgctgagccaggtgaggaagcagcagcagcagcagcagcttccc	1440
DB	1381	agccacatgtgctgctgctgagccaggtgaggaagcagcagcagcagcagcagcttccc	1440
QY	1441	caagctggtgagcagatgactcctggagaggtttggaaggtgctcctgctcagattctcc	1500
DB	1441	caagctggtgagcagatgactcctggagaggtttggaaggtgctcctgctcagattctcc	1500
QY	1501	acatcagttcaagtccaagggtgcaaaatgagtagcgtgcaagcgtgtaaggggaa	1560
DB	1501	acatcagttcaagtccaagggtgcaaaatgagtagcgtgcaagcgtgtaaggggaa	1560
QY	1561	tgctgtggaagcctcagcctgggaatctgcacagaggaagcgtggtgcccaggagga	1620
DB	1561	tgctgtggaagcctcagcctgggaatctgcacagaggaagcgtggtgcccaggagga	1620
QY	1621	gcagagggcgcgcacatcatggtgcccaggtgcagctgtgaggtcgtgactgctgactg	1680
DB	1621	gcagagggcgcgcacatcatggtgcccaggtgcagctgtgaggtcgtgactgctgactg	1680
QY	1681	gcagcctactcaaggtatcctaaagtgcagcagcagcagcagcagcagcagcagcagc	1740
DB	1681	gcagcctactcaaggtatcctaaagtgcagcagcagcagcagcagcagcagcagcagc	1740
QY	1741	tgccctcctgctcctgggagggcctcctggtgggactcctgctcctgcataaaaggaggt	1800
DB	1741	tgccctcctgctcctgggagggcctcctggtgggactcctgctcctgcataaaaggaggt	1800
QY	1801	gattttactgtgttattgtgttcttcaaatgtcttagtagtacctcattcaag	1860
DB	1801	gattttactgtgttattgtgttcttcaaatgtcttagtagtacctcattcaag	1860

[illegible]

RESULT

RESULTS 2
AAZ88157
ID AAZ88157 standard; cDNA; 1861 BP.

DT 25-APR-2000 (first entry)

XX	DE	Mouse testis specific factor	tesmin encoding	CDNA SEQ	ID NO:2.

Testis specific factor; tesmin; cell death; regulation; spermatocyte;
differentiation regulatory factor; male germ cell regulatory actor;
germ cell differentiation; sterility; ss.

Mus musculus.

Key	Location/Qualifiers
CDS	271..1158
FT	/*tag= a
FT	/product= "tesm10"

PN WO200004147-A1.

27-JAN-2000

16-JUL-1999: 99WO-JP03859.

PR 17-JUL-1998; 98JP-0219856.

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

DR WPI; 2000-147785/13.
DR P-PSDB; AAY68463.

AA New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility -

PS Example 2; page 38-42; 63pp; Japanese.

CC The present sequence encodes a male germ cell regulatory factor expressed specifically in spermatocytes, designated tesmin. Tesmin

Qy 1455 cagatgacccctggaggagtttggaaggtgctgcagattctccacatcagattcaag 1514
 |||||
 Db 1075 cagatgacccctggaggagtttggaaggtgctgcagattctccacatcagattcaag 1134
 |||||
 Qy 1515 tccaagggtgaaaaattgagtagctgcaagctggtaaaagggaatgcctgtggcaagc 1574
 |||||
 Db 1135 tccaagggtgaaaaattgagtagctgcaagctggtaaaagggaatgcctgtggcaagc 1194
 |||||
 Qy 1575 ctcaagccctggaaatctgcaccaggaagctggtgccaggagagagagagcgcgca 1634
 |||||
 Db 1195 ctcaagccctggaaatctgcaccaggaagctggtgccaggagagagagcgcgca 1254
 |||||
 Qy 1635 tcatggcaggtcagctgaggtctgagtgatctgcatggtactggtccagcctactcaa 1694
 |||||
 Db 1255 tcatggcaggtcagctgaggtctgagtgatctgcatggtactggtccagcctactcaa 1314
 |||||
 Qy 1695 ggtatctaaagtcaagcaggcgagagccacccctgggagtgagacactggccctctctcc 1754
 |||||
 Db 1315 ggtatctaaagtcaagcaggcgagagccacccctgggagtgagacactggccctctctcc 1374
 |||||
 Qy 1755 ctggggagccctctgggagcctccctccctgcataaaagagaggtgatttctactgt 1814
 |||||
 Db 1375 ctggggagccctctgggagcctccctccctgcataaaagagaggtgatttctactgt 1434
 |||||
 Qy 1815 tgttatgtgttcttcaaaattgcttagtagtacctccctccattcaagtattatgagccag 1874
 |||||
 Db 1435 tgttatgtgttcttcaaaattgcttagtagtacctccctccattcaagtattatgagccag 1494
 |||||
 Qy 1875 cctcaagtagagagtaggctctctctcaggtgagctgcgccaaatcaatacaagtc 1934
 |||||
 Db 1495 cctcaagtagagagtaggctctctctcaggtgagctgcgccaaatcaatacaagtc 1554
 |||||
 Qy 1935 agtggccatcaggggtttttccaggcagcctgtgacaggagatatggagggggtc 1994
 |||||
 Db 1555 agtggccatcaggggtttttccaggcagcctgtgacaggagatatggagggggtc 1614
 |||||
 Qy 1995 ggggttagagctgggtttgttggtatttttgcgttttttctctctgatttctgcttga 2054
 |||||
 Db 1615 ggggttagagctgggtttgttggtatttttgcgttttttctctctgatttctgcttga 1674
 |||||
 Qy 2055 agtgagaaaaactgtctctctgcacaccccttttccataaattactgtgcagctgcct 2114
 |||||
 Db 1675 agtgagaaaaactgtctctctgcacaccccttttccataaattactgtgcagctgcct 1734
 |||||
 Qy 2115 gctgaccagtcacagtgacctcagacaccagagaggtgaggttattatgcccacact 2174
 |||||
 Db 1735 gctgaccagtcacagtgacctcagacaccagagaggtgaggttattatgcccacact 1794
 |||||
 Qy 2175 ttgtgtttgtgtgagataaaactttccagactcccaaaaaaataaaaaa 2234
 |||||
 Db 1795 ttgtgtttgtgtgagataaaactttccagactcccaaaaaaataaaaaa 1854
 |||||
 Qy 2235 aaaaaa 2241
 |||||
 Db 1855 aaaaaa 1861
 |||||

RESULT 3

AZ88156
 ID AZ88156 standard; cDNA; 2134 BP.

XX
 AC

XX AZ88156;

XX 25-APR-2000 (first entry)

XX Human testis specific factor tesmin encoding cDNA SEQ ID NO:3.

XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH 407..1306
 FT /*tag= a
 FT /product= "tesmin"
 XX WO200004147-A1.
 PN 27-JAN-2000.
 PD
 XX 16-JUL-1999; 99WO-JP03859.
 PF
 XX 17-JUL-1998; 98JP-0219856.
 PR
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 PI WPI; 2000-147785/13.
 XX P-PSDB; AAY68464.
 DR
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 PT Claim 3; Page 42-47; 63pp; Japanese.
 PS
 XX The present sequence encodes a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX Sequence 2134 BP; 586 A; 490 C; 528 G; 530 T; 0 other;
 SQ
 Query Match 28.9%; Score 647.8; DB 21; Length 2134;
 Best Local Similarity 76.1%; Pred. No. 1.3e-169;
 Matches 829; Conservative 0; Mismatches 247; Indels 13; Gaps 2;
 Qy 495 gtgaaatacaag-aacaggtgtagtgccagggcgagccctgaagacgagcttt 553
 |||||
 Db 250 gttgaaatacaaggtaagcaggtggtagctacacagtaataatccggaagaacattt 309
 |||||
 Qy 554 ccaggccctctgctcaggaatcctgttgcaggttcccatcctccagaggcagagga 613
 |||||
 Db 310 gcagaattcttgcctcaggaatcctgttgcaggttcccatggtcccggaactagagga 369
 |||||
 Qy 614 ggctccagctgcccctgggaagaaactccagcccctggtgattgtcagctgaaagg 673
 |||||
 Db 370 tgctctctgttctcttaagaagaattccaaacccaatggtgatgccaattgaaagg 429
 |||||
 Qy 674 agggcccaagatgctctgcataagaaactgtggcgagggagctcaaaagcgtcccatct 733
 |||||
 Db 430 gggcacacaaatgctatgtatagaaattctagaaagagaaactaaagcactccattt 489
 |||||
 Qy 734 gcttccctcagtagatgaccagagagcagtttccctcagtcagagctccctaaagcaatgac 793
 |||||
 Db 490 ggttccctcagtagatgaccagagagcagtttccctcagtcagagctccctaaagcaatgac 549
 |||||
 Qy 794 aactttagtggaagacttctgcccagtcacagcagaatttaaatctcatcacaggttga 853
 |||||
 Db 550 tgctttagtaggagatttttgcagcagatacaaaaaataaataatctcatcacacaaact 609
 |||||
 Qy 854 taatggagctctccatcagctgcaatgggctgctcccttccctctgacctgctctgca 913
 |||||
 Db 610 tgaggggagcttaccatcggtagccaacgggtctgttccctctcggtgacaaactctcc 669
 |||||
 Qy 914 agggcccaaaaaataaactctgtggtgtagctgtgactgtcttccagcggggactctg 973
 |||||
 Db 670 aggaccacaaaaataaacttggctggtagctgtgactgttccagtgagggactttg 729
 |||||
 Qy 974 caacagctgca-----gctgcaacaacactcgccatagagctgagcttcaa 1021
 |||||

Db 730 caacaactgcaattgtaataattgttgcaacaactgcatcatgatattgaacggttttaa 789
Qy 1022 agccataaagcgctgtctgtatagaaaacctgaagctttccaacaaaatgggaaagg 1081
Db 790 agccattaaaggcatgcttggcagaaaaccagaagctttccagcccaaaaatgggaaagg 849
Qy 1082 ccgtgtggagctgtcaacttcgacacagcaaaagggtgcaactgtaagcgctcaggctg 1141
Db 850 ccaattgggcaatgccaagcccgacacaaagggtgcaactgcaggaggtcaggctg 909
Qy 1142 cctgaagaactactgtagctgactgaagggccaaaatcatgtgttccatttgcaaatg 1201
Db 910 cctgaagaattactgcgagtgctatgagggccaaaatattgtgttcttatttgcaaatg 969
Qy 1202 cattgtctgcaaaaactatgaagaagtcacagacgaagaaatgctgatgacacacccca 1261
Db 970 cattgttgcataaattatgaagaagcccgaaagcaaacactaatgagcatgccaaa 1029
Qy 1262 ctacatgagcctggagctttgagagcagccattattgtcccgagccaaagtcttcagg 1321
Db 1030 ctacatgcagactggaggtttggaaggcagccattacctgccacacaaagaattttcagg 1089
Qy 1322 acctccaaaactgaaaaaataggcagggcctctcctgtatctcctctgggaagtgtgga 1381
Db 1090 acttccaaagattcagtcacgatagcgcgcttccctcatgcatctcctggaggtgtgga 1149
Qy 1382 ggcacacatgcctgcctgctgcccaggggtgagaagcagagcaggagcactgttcccc 1441
Db 1150 ggcacacagcgcctgctgctgctcagggaggaagagggccgagaagacactgctccaa 1209
Qy 1442 aagcttggctgagcagatgactcctgagaggtttggaaggtgcctgtcgcagattctcca 1501
Db 1210 gtgcctggcagagcagatgactcctgaggaatttggaaggtgcttatcacagattctcca 1269
Qy 1502 catcagttcaagtcacaggggctgaaatbgatgtagctgcaagctggttaagggggaat 1561
Db 1270 cactgagtttaactaaaggattgaaatggagtagagataaaagtgtgaatgatgtt 1329
Qy 1562 gcctgtgac 1570
Db 1330 gatattgtc 1338

RESULT 4
ID AAL01300 standard; cDNA; 445 BP.
XX
AC AAL01300;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen cDNA SEQ ID NO: 1301.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ss.
OS Homo sapiens.
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.

QY 1442 aagcttggtgagcagatgactctgagagagtttgaaagtgcctgctgcagagattctcca 1501
 Db 2470 ggcagcagcggaacgagatgactctgaggaatncggaacnattgtgatgtcatcaa 2529
 QY 1502 catcgagttcaagtccaa 1519
 Db 2530 ctctgcaggaagggcaaa 2547
 RESULT 7
 AAH35563
 ID AAH35563 standard; cDNA; 250 BP.
 XX
 AC AAH35563;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:2645.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 11; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX
 WPI; 2001-235357/24.
 DR P-PSDB; AAG76158.
 DR
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PT
 PS Claim 1; Page 4257-4258; 9803pp; English.
 XX
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB7789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 250 BP; 71 A; 41 C; 69 G; 64 T; 5 other;

Query Match 5.4%; Score 120.4; DB 22; Length 250;
 Best Local Similarity 76.3%; Pred. No. 2.7e-23;
 Matches 148; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1377 gtggaggccacatgtgctgctgctgctgcccaggtgaagaagcagagcagcactgt 1436
 Db 2 gtggaggccacatgtgctgctgctgctgcccaggtgaagaagcagagcagcactgt 61
 QY 1437 tccccaagcttgctgagcagatgatccctggagagtttggaagtgctgctgcagatt 1496
 Db 62 tccaagtgcctggcagagcagatgatccctggagagtttggaagtgctgctgcagatt 121
 QY 1497 ctcccatcgagtccaagtcgaagggtgaaaaattgagtagctgcaagctggtaagg 1556
 Db 122 ctcccatcgagtccaagtcgaagggtgaaaaattgagtagataaagtgtgaatgc 181
 QY 1557 ggaatgcctgtggc 1570
 Db 182 atgttgatttgc 195
 RESULT 8
 AAS82145
 ID AAS82145 standard; cDNA; 1093 BP.
 XX
 AC AAS82145;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #17949.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 WPI; 2001-639362/73.
 DR P-PSDB; ABG17958.
 DR
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 1; SEQ ID No 17949; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 1093 BP; 325 A; 185 C; 267 G; 316 T; 0 other;

Query Match 4.7%; Score 105.4; DB 23; Length 1093;
Best Local Similarity 65.4%; Pred. No. 9.2e-19;
Matches 176; Conservative 0; Mismatches 81; Indels 12; Gaps 1;

Qy 941 gtactgtgactgcttccagcggggacttctgcaacagctgca-----gctg 988
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Db 589 gtattgtgattgttgcgaatgtgaatttgcgaactgcaatgtactaattgta 648
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Qy 989 caacaactgcgcagctgagcgtctcaaaagccataaaagcgtgtcttgatagaaa 1048
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Db 649 caacaatttgaaatgaaatgaaagcgaagcaataaaagcgtgttgacagaaa 708
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Qy 1049 tcctgaagcttcccaacaaaatggggaagcgcgtctggagctgctaaactcgaca 1108
||| ||||| ||| ||| ||| ||||| ||||| ||||| |||
Db 709 tccagaagcctttaagcctaagatagggaaggaaggaaggaagcgtgacgtca 768
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Qy 1109 cagcaaaagggctgaactgttaagcgtcagcgtcgtgaaagaactactgtgagtgctatga 1168
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Db 769 tagcaaaagggctgaactgcaaacgactcaggtgcttcaaaactactgtgaatgctatga 828
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Qy 1169 ggcacaaatcatgtgtcttccatttga 1197
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||
Db 829 gaacattatttccatttggattgttga 857
||| ||||| ||||| ||| ||| ||| ||||| ||||| |||

RESULT 9

ABL10437
ID ABL10437 standard; cDNA; 3164 BP.

XX ABL10437;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 25793.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66334.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Claim 1; SEQ ID NO 25793; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 3164 BP; 880 A; 865 C; 810 G; 609 T; 0 other;

Query Match 4.1%; Score 92.8; DB 23; Length 3164;
Best Local Similarity 59.5%; Pred. No. 5.4e-15;
Matches 176; Conservative 0; Mismatches 117; Indels 3; Gaps 1;

Qy 949 actgtctctccagcggggacttctgcaacagctgcaacaaactgcgcctgagc 1008
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Db 2584 acggcgaatttgcaggactgcactgcaaggattctttaacaactctgactacgaag 2643
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Qy 1009 tcgagcgtctcaaaagccataaaagcgtgctctgatagaatcctgaagcttccaaacaa 1068
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Db 2644 tggagcggggcgtgctattcgacgtcgtcgtatcgttaacccagcgtcttaaaccca 2703
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Qy 1069 aaatggggaaagcgcgtctggagctgctaaacttcgacacagcaaaaggtgcaactgta 1128
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Db 2704 aaattacggcaccat---tcaggtgatatgctgctgatacaacaaggttgcaactgca 2760
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Qy 1129 agcgtcagcgtcgtgaaagaactactgtgagtgctatgagcccaaaatcatgtgtcttt 1188
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Db 2761 aaagatcggcgtcctcaagaactattgtgagtgctatgagcaaaagattccctgctcca 2820
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Qy 1189 ccattgcaaatgctgttcgaaactatgaagaaagtcagaaagccgaaacaaatg 1244
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

Db 2821 gcatagtaaatcgttgggtgtcgaaacatggaagccgtccggagcgtggacatg 2876
||| ||| ||| ||||| ||| ||| ||| ||||| ||||| |||

RESULT 10

AAV72865

ID AAV72865 standard; cDNA; 1503 BP.

XX AAV72865;

XX 29-MAR-1999 (first entry)

XX Caenorhabditis elegans synMuv gene lin-54 cDNA.

XX LIN-54; synthetic multivulvar; SynMuv; signal transduction;

KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;

KW cell proliferation; gene therapy; ss.

XX Caenorhabditis elegans.

XX Key Location/Qualifiers

FT CDS 17..1333

FT /*tag= a

XX WO9854299-A1.

XX 03-DEC-1998.

XX 28-MAY-1998; 98WO-US11043.

XX 28-MAY-1997; 97US-0047996.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Ceol C, Horvitz HR, Lu X;

XX WPI; 1999-045362/04.

XX P-PSDB; AAW83392.

XX Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
PT useful for treating diseases associated with altered levels of cell

Duan RD, Florence KA, Soppet DR;
WPI; 2000-679828/66.
P-PSDB; AAB56124.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -

Claim 1; Page 882; 1065pp; English.

The polynucleotide sequences given in AAC99818 to AAC99977 encode the human secreted proteins given in AAB56077 to AAB56362. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiantic; vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial; virucide; fungicide; and opthalmological. The human secreted polynucleotides and proteins can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The proteins can also be used as a food additive or preservative to increase or decrease storage capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used in the exemplification of the present invention.

Sequence 372 BP; 130 A; 73 C; 86 G; 78 T; 5 other;

Query Match 2.0%; Score 45.6; DB 21; Length 372;
Best Local Similarity 68.5%; Pred. No. 0.022;
Matches 63; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 2150 tgagggtgcttattatgccacacttggttttgtgtgagataaacctttccagact 2209
 ||| || | | | | | | | | | | | | | | | | | | |
Db 263 tcatatgccttccttatcaccatgttggttgatgcaataaaagtgtttctctt 322

QY 2210 cccaaaaaaaaaaaaaaaaaaaaaaaa 2241
 || ||||| ||||||| ||||||| |||||
Db 323 ccaaaaaaaaaaaaaaaaaaaaaaaa 354

RESULT 14
AAI89088
ID AAI89088 standard; cdNA; 387 BP.
XX
AC AAI89088;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 9148.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
OS Homo sapiens.
XX
PN W0200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-514838/56.
DR P-PSDB; AAO09157.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX Claim 1; SEQ ID NO 9148; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pt_sequences.
XX Sequence 387 BP; 127 A; 65 C; 63 G; 119 T; 13 other;

Query Match 2.0%; Score 45.2; DB 22; Length 387;
Best Local Similarity 66.3%; Pred. No. 0.03;
Matches 65; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 2144 agaaggtgagtggtcttatgcccacacttctgtgtgtgtgtgagataaaccttcc 2203
Db 110 agttgttagaagatgggtattgcatgtcactttttttgtgataaaacacttcc 169
QY 2204 cagactcccaaaaaaaaaaaaaaaaaaaaaaa 2241
Db 170 ctcccaaaaaaaaaaaaaaaaaaaaaaa 207

RESULT 15
AAFI6214
ID AAFI6214 standard; cDNA; 1696 BP.

XX AAFI6214;

XX 13-MAR-2001 (first entry)

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:649.

XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.

XX Homo sapiens.

XX WO200055174-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05988.

XX

PR 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM;
PI WPI: 2000-587513/55.
DR P-PSDB; AAB57011.
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX Claim 1; Page 1090; 2338pp; English.

XX AAFI5566 to AAFI6505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAFI6506 to AAFI6514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.

XX Sequence 1696 BP; 452 A; 465 C; 430 G; 347 T; 2 other;

Query Match 2.0%; Score 44.8; DB 21; Length 1696;
Best Local Similarity 81.2%; Pred. No. 0.087;
Matches 52; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2178 tgtttgtgtgagaataaaccttccagactcccaaaaaaaaaaaaaaaaaaaaaa 2237
Db 1611 tgttttaaatgacaataaaacagccagaatgtcaaaaaaaaaaaaaaaaaaaaaa 1670

QY 2238 aaaa 2241

Db 1671 aaaa 1674

Search completed: July 11, 2002, 10:54:07
Job time: 9564 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:14:43 ; Search time 3455.78 Seconds
(without alignments)
8752.489 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 2241
Sequence: 1 tatcctgtggttgcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_estl:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1291.4	57.6	1730	11 AK015724	Mus muscu
c	647.2	28.9	708	10 BG076362	H3158G12-
2	604.8	27.0	631	9 AV282581	AV282581
3	505.8	22.6	544	10 BG101151	uy20a06.y
4	499.2	22.3	545	12 AZ661404	1M0540B04
5	493.2	22.0	505	10 BG088886	H3158G12-
6	449.2	20.0	2010	11 AK015732	Mus muscu
c	440.2	19.6	468	9 AW558944	AW558944
7	379.8	15.3	418	10 BF452508	uz97g12.y
8	341.8	16.9	842	10 B1829802	603079876
c	333	14.9	374	10 BF152725	uz95a04.x
9	314.8	14.0	329	12 AZ660217	1M0538L08
10	265.4	11.9	768	10 BG977029	602845588
11	227.4	10.1	276	9 B328330	BB328330
12	224.4	10.0	247	9 AV254566	AV254566
13	221	9.9	275	9 AV260483	AV260483
14	202.4	9.0	654	9 BB617220	BB617220

18	194.6	8.7	244	9	AV256079	AV256079
19	184.6	8.2	241	9	AV045606	AV045606
20	175.6	7.8	283	9	BB068350	BB068350
21	173	7.7	697	9	BB611478	BB611478
22	171.8	7.7	312	9	AV277921	AV277921
23	171.4	7.6	574	9	AI981460	pat.pk005
24	171	7.6	333	9	AL040101	DKF2P434C
25	167.4	7.5	221	9	BB536930	BB536930
26	166.8	7.4	257	9	AV265406	AV265406
c	166	7.4	420	10	BG272409	nanh30a05.
28	164.2	7.3	475	10	BI706999	fq13f03.y
29	159.8	7.1	266	9	AF012383	AF012383
30	149.4	6.7	397	10	BF937310	fm3c07.y
31	149	6.6	704	9	AV715308	AV715308
32	145.6	6.5	176	9	AV261053	AV261053
33	138.8	6.2	204	9	AV343095	AV343095
34	133.8	6.0	241	9	AV282159	AV282159
35	133.6	6.0	638	10	BF426275	df69g02.y
36	124.8	5.6	285	9	AV091958	AV091958
c	123	5.5	696	12	AZ512200	1M0357P04
38	120	5.4	312	9	AW270088	xv37g12.x
39	111.6	5.0	118	10	BF152726	uz95a04.y
c	107.4	4.8	442	9	AI307797	tb28c05.x
41	103	4.6	918	10	BG916417	602814015
42	95.2	4.2	888	10	BG773009	602721268
c	94	4.2	642	9	AI968415	wu02b05.x
44	86.6	3.9	271	10	BG184200	RST3121.A
45	74.4	3.3	552	10	B1421583	EST532249

ALIGNMENTS

RESULT 1

AK015724	AK015724	1730 bp	mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930507A18.metallothionein-like 5, testis-specific (tesmin), full insert sequence.	linear	HTC 19-JAN-2002
LOCUS					
DEFINITION					
ACCESSION	AK015724				
VERSION	AK015724.1	GI:12854168			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus				
REFERENCE	1 (sites)				
AUTHORS	Carninci,P. and Hayashizaki,Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2 (sites)				
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				

Db 1498 CTCAGCCCTGGAAATCTGCACCGAGGAAGTGGTGCCCGAGGAGCAGAGGCGCGCA 1557
 Qy 1635 teatggcaggtcagctgaggtctgagtgatctgactgctgactgagccactactca 1694
 Db 1598 TCATGGCCAGGTTCAGCTGAGGCTGAGTGTGATCTGCTGCTGCTGCTGCTGCTGCT 1617
 Qy 1695 ggtatccctaaagtcaaacgagcagagccacccctgggagtgagacactgcccctctgccc 1754
 Db 1618 GGTATCCTTAAGTGAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1677
 Qy 1755 ctggggagggccctctggggagctccctgcccctgcataaaagagggtgattttc 1807
 Db 1678 CTGGGGAGGCCCTCTGGGAGCTCCCTGCGCTGCATATAAAGAGGGTGATTTC 1730

RESULT 2
 BG076362/c
 LOCUS
 DEFINITION H3158G12-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 H3158G12 3', mRNA sequence.
 ACCESSION BG076362
 VERSION BG076362.1 GI:12558931
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 708)
 Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka
 T.S., Carter, M.G. and Ko, M.S.H.
 Verification and initial annotation of NIA mouse 15K cDNA clone set
 Unpublished (2001)
 Other ESTs: H3158G12-5
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit <http://lgsun.grc.nia.nih.gov/cdna/15k.html> for details.
 Plate: H3158 row: G column: 12
 Seq primer: -21M13 Forward
 High quality sequence stop: 708
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1..708
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="niaEST:H3158G12-3"
 /db_xref="taxon:10090"
 /clone="H3158G12"
 /clone_lib="NIA Mouse 15K cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"
 /note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryos cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dt)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of

7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
 Mol Genet 7: 1967-1978."
 BASE COUNT 188 a 197 c 171 g 152 t
 ORIGIN

Query Match 28.9%; Score 647.2; DB 10; Length 708;
 Best Local Similarity 98.9%; Pred. No. 1.2e-96;
 Matches 704; Conservative 0; Mismatches 3; Indels 5; Gaps 5;
 Qy 1502 catcgagttcaagtcacaggggctgaaaattgagtagctgctgagtgatctgcatggtactgg 1561
 Db 708 CATCGAGTTCAAGTCCAA-GGCGCTGAAATTTAGTAGCTGCAAGCTGGTAAAGGGGAAT 650
 Qy 1562 gctgtggcaagcctcagccctgggaattctgcacccagaggaagctggtgccacggagagag 1621
 Db 649 -CCTGTGCAAGCCCTCAGCCCTGGGAATCTGCACCGAGGAAGCT-GTCCCGAGGAGGAG 592
 Qy 1622 cagagggcgcgcacatcagccaggtcagctgaggtctgagtgatctgcatggtactgg 1681
 Db 591 CAGAGGCCCGGCATCATGGCCAGGTCAAGCTGTGAGGTCTGAGTGTATCTGCTGCTGCTG 532
 Qy 1682 ccagcctactcaaggtatcctaaagtgcacagcagagccacccctgggagtgagacact 1741
 Db 531 CCAGCCTACTCAAGGTATCCTTAAAGTGCAAGCAGCAGCAGCAGCAGCCTGGGATGCACT 472
 Qy 1742 ggcctcctctccctggggagccctctggggactccctgcctgcataaaagagggtg 1801
 Db 471 GGCCTCTCTCTCTGGGAGGCCCTCTGGGAGCTCCCTGCTGCTGCATATAAAGAGGGTG 412
 Qy 1802 atttctactgtgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1861
 Db 411 ATTTTCTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 352
 Qy 1862 tattatgagccagcctcaagtttagagagctgaggtctctctcagctgaggtcgtcccaaa 1921
 Db 351 TATTATGAGCCAGCCCTCAAGTTAGAGAGCTAGGCTCTTCTTCAGGTGGACTCTGCCCCAA 292
 Qy 1922 tcaatacaagtcaggtggccatcaggggttttccagggcagggcctgtgacaggagata 1981
 Db 291 TCACATAGAAGTCAGGTGGCCATCAGGGG-TCTTCCAGGCCAGGCTGTGACAGGAGATA 233
 Qy 1982 tggaggggggtcgggttagagctggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2041
 Db 232 TGGGAGGGGGTTCGGGTTAGAGCTGGGTTGTGTGGATTTTGTGGATTTTGTGGATTTTGT 173
 Qy 2042 tattctgtcgtgaagtgcagaaactgtctcctgtccaa-cctttctccataataactg 2100
 Db 172 TATTTCTGCTTGAAGTGAGAAAACTTGTCTCTGTCCAACCCCTCTTCCATAAATACTG 113
 Qy 2101 ctgacaggtcgcctcgtcagccagtcacagtcacagtcacagtcacagtcacagtcacagtcac 2160
 Db 112 CTGACAGGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53
 Qy 2161 attatgccacactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2212
 Db 52 ATTATGCCACACATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1
 RESULT 3
 LOCUS AV282581
 DEFINITION AV282581 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4933430L08 3', mRNA sequence.
 ACCESSION AV282581
 VERSION AV282581.2 GI:16391671
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 631)

E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1420587
Seq primer: Primer name ambiguous
High quality sequence stop: 359.

FEATURES
source
1..544
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:3658283"
/clone_lib="McCarrey Eddy spermatocytes"
/sex="male"
/tissue_type="spermatocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pBluescript SK+ (Stratagene)
); Site_1: XhoI; Site_2: EcoRI; CDNA oligo dt-primed
[5'-(GA)10-AC TAGCTCGAGTTT-3'] and directionally
cloned using 5' linkers 5'-AATTCGGCAGAG-3' and
5'-CTCGTGGCG-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and transformed
into DH10B. Library contains 98% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."

BASE COUNT 123 a 134 c 164 g 122 t
ORIGIN

Query Match 22.6%; Score 505.8; DB 10; Length 544;
Best Local Similarity 96.9%; Pred. No. 1.9e-73;
Matches 527; Conservative 0; Mismatches 13; Indels 4; Gaps 1;

QY 1338 aaaataggcagccttctctgtatctctgtggaagtagtgagggcacatgtgcctgc 1397
DB 1 AAAAATAGGCAGCCCTTCTCTGTATCTCTGGGAAGTAGTGGAGGCCACATGTCCTGC 60

QY 1398 ctgctggccagggtaggaagcagcagaggagcactgttcccccaagcttggctgagcag 1457
DB 61 CTGCTGGCCAGGGTGAGGAAGCAGCAGAGGACACTGTTCCTCCCAAGCTTGGCTGAGCAG 120

QY 1458 atgatctggagagatttggaaggctgctgtcgagattctccacatcgagttcaagtc 1517
DB 121 ATGATCTGTGGAGAGGTTTGGAAAGGTGCTGTGCGAGATTCTCCATCGAGTTCAGTGC 180

QY 1518 aaggggctgaaattc---gagtagcgtgcagctggtgaaagggaatgcctgtgagcag 1573
DB 181 AAGGGGCTGAAAAATGAATACCGGTGCAACCGTGGTAAGGGGAATGCCCTGTGGCAG 240

QY 1574 cctcagccttgggaattctgcaccggaggagctgggtgccaggaggagcagagggcgccg 1633
DB 241 CCTCAGCCTGGGAATCTGCACCGAGGAGCTGGTGGCCAGGAGGAGCAGAGCGCGC 300

QY 1634 atcatggccaggtcaggtctgaggtctcagtgatctcagtgatctcagtcagtcagtc 1693
DB 301 ATCATGGCCAGGTCAGCTGTGAGGTCTCAGTGATCTGATGTCATGTCGTGTCGCGCCTACTCA 360

QY 1694 eggtatctctaaagtcaagcagcagcagccctgggagtagcacactgcccctctgtc 1753
DB 361 AGGTATCTCTAAAGTGCAGCAGCAGAGCCCTGNGGATGACACTGCGCTCTGTGTC 420

QY 1754 cctggggaggccctctggggactccctgcctgcataaaagagggtgattttctacttg 1813
DB 421 CCTGGGAGGCCCTCTGGGGACTCCTCGCTGCATAAAAGAGGGTGATTTTCTACTTG 480

QY 1814 ttgttatgtttgttttcaaatctttagtagtacctccattcaagtattatgatgcca 1873
DB 481 TTGTTATGTGTTTGTCTTCAAAATGCTTAGTAGTACCTCCATTCAGATTATTATGAGCCA 540

QY 1874 gcct 1877
DB 541 GCCT 544

RESULT 5
LOCUS AZ661404 545 bp DNA linear GSS 14-DEC-2000
DEFINITION 1M0540B04F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0540B04 F, DNA sequence.
ACCESSION AZ661404
VERSION AZ661404.1 GI:11798550
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 545)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0540 row: B column: 04
Seq primer: CGTTGTAACACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 545.
Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0540B04"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi14732114|gb|AF129072.1), a copy-number

Qy 933 ctgtctgggtactgtgactcttctccagcgagggaacttctgcaacagctcagctgcaac 992
 |||||||
 Db 361 CTGTCTGGGTACTGTGACTGCTTCTCCAGCGGAGACTTCTCAACAGCTGCAGCTGCAAC 420
 |||||||
 Qy 993 aacctgcgcctagctcagctcagctcgaacccataaa-gcgtgtcttgcatagaataacc 1051
 |||||||
 Db 421 AACCTGCGCCATGAGCTCGAGCGCTTCAAGCCATAAAGGCGTGTCTTGATAGAAATCC 480
 |||||||
 Qy 1052 tgaagctttccaacaaataatgggg 1076
 |||||||
 Db 481 TGAAGCTTTTCCAAACCAAAATGGGG 505
 |||||||

RESULT 7
 AK015732 2010 bp mRNA linear HTC 19-JAN-2002
 LOCUS AK015732
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930509C02:metallothionein-like 5, testis-specific (tesmin), full insert sequence.
 ACCESSION AK015732
 VERSION AK015732.1 GI:128541177
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:4930509C02.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 92279253
 PUBMED 10349636
 REFERENCE 2 (sites)
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3 (sites)
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaiguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4 (sites)
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5 (bases 1 to 2010)
 AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schraml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamana,K., I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATCCCTCCCTCCCTCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
 Location/Qualifiers
 source 1. 2010
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGD:MGI:1907904"
 /db_xref="taxon:10090"
 /clone="4930509C02"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 gene 67. 900
 /gene="Mt15"
 CDS 67. 900
 /gene="Mt15"
 /note="data source:MGD, source key:MGI:1340029, evidence:ISS
 metallothionein-like 5, testis-specific (tesmin)
 putative"
 /codon_start=1
 /protein_id="BAB29949.1"
 /db_xref="GI:12854178"
 /translation="MEDALLGAMTGPEDLFGSERVFADGLALSPAGGAADRD
 LPVIADAYLGATEPEPLRLALSPPEAALGDFPGLPEURPDDRAAPPAYS
 HVSSLLPARGPALPLSAGVRVIVPEIKAGGVPGGPDAFAAPLAQESCKRF
 PSSQEAEEASSCKDPSPMVICQLGGAQMLCIDNGARELKHLLHLLPOYDQSSF
 POSELKPMTTLVGRLLPVPKALNLTQVDNGALPSAVNGAFAFSGPALQGPRTILS
 G"
 polyA_signal 1991. 1996
 /note="putative"
 polyA_site 2010
 /note="putative"
 BASE COUNT 407 a 558 c 563 g 482 t
 ORIGIN
 Query Match 20.0%; Score 449; DB 11; Length 2010;
 Best Local Similarity 100.0%; Pred. No. 2.le-64;
 Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 495 gttgaatacaagaagcagggtggtgtgtagtgccaggcgccctgaagacagcagctttc 554
 |||||||
 Db 451 GTTGAATCAAGAAGCAGCGTGGTAGTGCCAGCGCGCCCTGAAGACCGCAGCTTTC 510
 |||||||
 Qy 555 caggccctctggctcaggaatcctctgtgcaagttcccatcatccccaggagcagaggag 614

|||||
 Db 511 CAGGCCCTCTGGCTCAGGAATCCTGTTCAAGTTCCCATCATCCAGAGGAGGAG 570
 QY 615 gctccagctgcccctggaagaagactccagcccatggtgattgtcagctgaaagga 674
 Db 571 GCCTCAGCTGCCCTCGGAAGAAGACTCCAGCCCCATGGTATTGTCAGTGAAGA 630
 QY 675 ggcgcagatgctctgcatagacaactgtgcgagagagctcaaaagcgtccatcg 734
 Db 631 GCGGCCAGATGCTGTGATAGACAACCTGTGGCGGAGGAGCTCAAGCGCTCCATCTG 690
 QY 735 ctctccagtaoatgacagagagcagttccctccagtcagagagctccctaaagcaatgaca 794
 Db 691 CTTCCTCAGTACGATGACAGAGCAGTCTCCCTCAGTCAGAGCTCCCTAAGCCAATGACA 750
 QY 795 actttagtggaagactctgcagaccaggaaggttaaatctcatcacacaggtgat 854
 Db 751 ACTTTAGTGGGAAGACTCTGCCAGTACCAGCGAAGTAAATCTCATCACACAGGTTGAT 810
 QY 855 aatggagctctcccatcagctgcaatggggtgccttccctctggacctgctgtgcaa 914
 Db 811 AATGGAGCTCTCCATCAGCTGTCAATGGGCTGCCTTTCCCTCTGGACCTGCTGTGCAA 870
 QY 915 gggccaccacaaataactctgtctgggta 943
 Db 871 GGGCCACCAAAATAACTCTGTCTGGGTA 899

RESULT 8
 AW558944/c
 LOCUS
 DEFINITION L0302B03-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
 clone L0302B03 3', mRNA sequence.
 ACCESSION AW558944
 VERSION AW558944.1 GI:7204373
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
 M.J., Pantano,S., Sano,Y., Piao,Y., Nagara,R., Doi,H., Wood,W.H.
 1 (bases 1 to 468)
 III, Becker,K.G. and Ko,M.S.H.
 Genome-wide expression profiling of mid-gestation placenta and
 embryo using a 15,000 mouse developmental cDNA microarray
 Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
 20381348
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: L0302 row: B column: 03
 Seq primer: -21M13 Forward
 High quality sequence stop: 468
 POLYA=Yes.

FEATURES
 source
 1. 468
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="niaEST:L0302B03-3"
 /db_xref="taxon:10090"
 /clone="L0302B03"
 /clone_lib="NIA Mouse Newborn Ovary cDNA Library"
 /sex="female"
 /dev_stage="Newborn Ovary"
 /lab_host="DH10B"
 /note="Vector: pSPORT1 (Gibco/BRL Life Technology);
 Site 1: SalI; Site 2: NotI; Total RNAs were extracted from
 7 Newborn Ovary. The double-stranded cDNA was synthesized
 by Gibco's kit with an Oligo(dT) primer [NotI
 primer-adaptor from GibcoBRL]

RESULT 9
 BF452508
 LOCUS
 DEFINITION uz97g12.y1 NCI_CGAP_Tel1 Mus musculus cDNA clone IMAGE:3809495 5',
 mRNA sequence.
 ACCESSION BF452508
 VERSION BF452508.1 GI:11518677
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 418)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

[5'-pCACTAGTCTAGATCGGAGCGGCCCTTTTCTTTTCTTTT-3'] from
 2.56ug of total RNA. The double-stranded cDNAs were
 treated with T4 DNA polymerase and purified by
 ethanol-precipitation. The cDNAs were ligated to
 Lona-linker LL-Sal3 (include SalI sequence). The cDNAs
 were purified by phenol/chloroform and separated from
 free linkers by Centricon 100. Then, cDNAs were amplified
 by long-range high fidelity PCR using Takara's Ex Taq
 polymerase. Then, the cDNAs were purified by
 phenol/chloroform and by Centricon 100. The cDNAs were
 digested with SalI and NotI enzymes. Then, the cDNAs were
 size selected by Gibco's Size Fractionation Column. The
 cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
 vector. The DH10B E. coli host was transformed with the
 ligation mixture by chemical method. The library was
 constructed by Xiaohong Wang and Yulan Piao."

BASE COUNT 146 a 117 c 112 g 93 t
 ORIGIN
 Query Match 19.6%; Score 440.2; DB 9; Length 468;
 Best Local Similarity 98.9%; Pred. No. 1.2e-62;
 Matches 464; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
 QY 1745 cctctgtccctggggagggccctctgggagctccctgcctgcataaaagaggtgatt 1804
 Db 468 CCTCTGTCTCTGGGAGGCGCTCTGGGACTCCCTGCATATAAAGAGGGTGATT 409
 QY 1805 ttctactgtgtatgtgttgccttcaaatgcttagtagtacctcattcaagttat 1864
 Db 408 TTCTACTTGTGTATGTGTGCTTTCAATGTGCTTAGTAGTACCTCCATTCCAGTTAT 349
 QY 1865 tatgagccagcctcaagtttagagagctgcttcttcaggctggactctgcccataatca 1924
 Db 348 TATGAGCCAGCCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGGACTCTGCCCAATCA 289
 QY 1925 catacaagtcaggtggccatcaggggttttccagggccagggcctgacagagatatgg 1984
 Db 288 CATACAAGTCAGGTGGCCATCAGGGG-TCTTCCAGGCCAGGCTGTGACAGGAGATATGG 230
 QY 1985 gaggggggtcggttagagctgggtttgttgatttttgcgttttttcttcctgtat 2044
 Db 229 GAGGGGGTTCGGGTAGAGCTGGGTTGTTGGATTGTTTTCCTTCTTCTTCTGAT 170
 QY 2045 ttctgctgaagtggagaaactgtctcctgtccaa-cctttctccataattactgctg 2103
 Db 169 TTCTGCTTGAAGTGAGAAAACTTGTCTCCTGTCCAACTCTTCTCCATAAATTACTGCTG 110
 QY 2104 caggtgcgctgctgaccagtcacagtcacccagacacagagtgaggtgcttatt 2163
 Db 109 CAGGGTGGCTGTGTGACCATCAGTACCTCAGACACCCAGAGGTGAGGTGGCTTATT 50
 QY 2164 atgcccacactttgtgttttgtgtgagataaaacctttccagactccc 2212
 Db 49 ATGCCACACACTTGTGTTGTTGTGAGAATAAACCTTTTCCAGACTCCC 1

Db	421	CTCATGCATCTCTGGAGGTGTGGAGGCCATCGCCTGCTGCTTGGTCTCAGGAG	480
Qy	1414	aggagcagagcagagcagcagctgtccccaaagcttggctgagcagatgatctggagaggt	1473
Db	481	AAGAGGCGGAGAAAGACACTGCTCCAAGTGCCTGGCAGACAGATCATCTCTGAGGAAT	540
Qy	1474	ttggaag	1480
Db	541	TTGGAAG	547
RESULT	11		
LOCUS	BF152725/c		
DEFINITION	BF152725 374 bp mRNA linear EST 26-OCT-2000		
ACCESSION	u295a04.x1 NCI_CGAP_Tel Mus musculus cDNA clone IMAGE:3809167 3'		
VERSION	similar to TR:Q9WTJ6 Q9WTJ6 TESMIN-2. ;, mRNA sequence.		
KEYWORDS	BF152725		
SOURCE	BF152725.1 GI:11034130		
ORGANISM	EST.		
REFERENCE	house mouse.		
AUTHORS	Mus musculus		
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
COMMENT	1. (bases 1 to 374)		
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
	Unpublished (1997)		
	Other_ESTs: u295a04.y1		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nlm.gov		
	Tissue Procurement: Jeffrey E. Green, M.D.		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
FEATURES	NCI:1451279		
source	Seq primer: -40UP from Gibco		
	High quality sequence stop: 316.		
	Location/Qualifiers		
	1. .374		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:3809167"		
	/clone_lib="NCI_CGAP_Tel"		
	/sex="male"		
	/dev_stage="5 months"		
	/lab_host="DH10B (T1 phage-resistant)"		
	/note="Organ: testis; Vector: PCMV-SPORT6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.		
	Average insert size 1.3 kb. Library constructed by Life		
	Technologies."		
BASE COUNT	75 a	115 c	84 g 100 t
ORIGIN			
Query Match	14.9%; Score 333; DB 10; Length 374;		
Best Local Similarity	97.0%; Pred. No. 4.7e-45;		
Matches 350;	Conservative	0; Mismatches 10; Indels 1; Gaps 1;	
Qy	1449	gctgagcagatgatcctggaggagcttggagggtgcctgtgcgagattctccacatcgag	1508
Db	374	GCTAAGAAAAATGATCCTCGGAGGAGTTTGGAGGTGCCTGCGCAGATTCTCCACATCGAG	315
Qy	1509	tccaagtccaagggtcgaaattgactacgtgcagctggtaaacggaatccctgtg	1568
Db	314	TTCAGTCCAAAGGGGCTGAAAATTTAGTAGCGT-TAAGCTGGTAAAGGGGAATGCCCTGTG	256
Qy	1569	qcaagcctcagcctcgaggaattctgcaccgaggaagctgggtgccccagggagagcagagcc	1628

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:14:43 ; Search time 95.08 seconds
(without alignments)
5789.490 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 2241

Sequence: 1 tatcctgtggttgcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	44.2	2.0	615	4	US-09-328-111-371
C 2	43.6	1.9	964	1	US-08-448-606-5
C 3	43	1.9	658	4	US-09-385-982-327
C 4	42.8	1.9	3024	6	5284931-1
C 5	41	1.8	1129	4	US-09-227-357-40
C 6	40.4	1.8	333	3	US-08-946-026-35
C 7	40.4	1.8	578	4	US-09-328-111-757
C 8	40.4	1.8	1977	4	US-09-227-357-83
C 9	40.4	1.8	2238	1	US-08-742-011-1
C 10	40.4	1.8	7218	1	US-08-232-463-14
C 11	40.2	1.8	672	1	US-07-847-010-21
C 12	40.2	1.8	1181	1	US-08-181-271A-27
C 13	40.2	1.8	1181	1	US-08-449-315-27
C 14	40.2	1.8	1181	1	US-08-444-803-27
C 15	40.2	1.8	1181	1	US-08-449-043-27
C 16	40.2	1.8	1181	1	US-08-456-265A-27
C 17	40.2	1.8	1181	1	US-08-455-416-27
C 18	40.2	1.8	1181	1	US-08-455-244-27
C 19	40.2	1.8	1181	1	US-08-454-876-27
C 20	40.2	1.8	1181	2	US-08-457-364-27
C 21	40.2	1.8	1181	2	US-08-456-262-27
C 22	40.2	1.8	1181	2	US-08-456-240-27
C 23	40.2	1.8	1181	2	US-08-455-736-27
C 24	40.2	1.8	1181	2	US-08-971-217-27
C 25	40.2	1.8	1181	4	US-09-350-600-27
C 26	40.2	1.8	2672	1	US-08-703-947-1
C 27	39.4	1.8	903	5	PCT-US95-06406A-21

```

28 38.6 1.7 882 2 US-08-909-965C-9 Sequence 9, Appli
29 38.6 1.7 1733 3 US-09-073-569-1 Sequence 1, Appli
30 38.6 1.7 3217 4 US-09-232-200-64 Sequence 64, Appl
31 38.6 1.7 3217 4 US-09-232-137-64 Sequence 64, Appl
32 38.6 1.7 3217 4 US-09-232-201-64 Sequence 64, Appl
33 38.6 1.7 3471 5 PCT-US93-00227-2 Sequence 2, Appli
34 38.4 1.7 770 3 US-08-938-675A-1 Sequence 1, Appli
35 38.4 1.7 1858 2 US-08-909-965C-11 Sequence 11, Appli
36 38.4 1.7 2409 4 US-09-293-322C-8 Sequence 8, Appli
37 38.4 1.7 152331 3 US-09-128-155-16 Sequence 16, Appli
38 38.2 1.7 658 4 US-09-328-111-816 Sequence 816, App
39 38.2 1.7 770 3 US-08-865-297-5 Sequence 5, Appli
40 38.2 1.7 2107 4 US-09-180-852-1 Sequence 1, Appli
41 38.2 1.7 2413 4 US-09-518-046-1 Sequence 1, Appli
42 38.2 1.7 2416 4 US-09-261-416-1 Sequence 1, Appli
43 38.2 1.7 2544 4 US-09-518-046-3 Sequence 3, Appli
44 38.2 1.7 3761 4 US-08-890-865A-2 Sequence 2, Appli
45 38 374 2 US-08-370-156-24 Sequence 24, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-328-111-371/c
; Sequence 371, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 371
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(615)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-371

```

Query Match 2.0%; Score 44.2; DB 4; Length 615;
Best Local Similarity 75.3%; Pred. No. 0.0055;
Matches 55; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 2169 cacacttggttgttgtagaataaacctttccagactcccaaaaaaaaaa 2228

Db 82 CAGACTTGTGTGTTTGTATTATTAAGCATTCCTTAACAGAAAAAAAAAAAA 23

Qy 2229 aaaaaaaaaaaaaa 2241

Db 22 AAAAAAAAAAAAAA 10

RESULT 5
US-09-227-357-40
; Sequence 40, Application US/09227357

RESULT 3
US-09-385-982-327/c


```

; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
; LENGTH: 1129
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1053)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-40

Query Match          1.8%; Score 41; DB 4; Length 1129;
Best Local Similarity 68.3%; Pred. No. 0.064;
Matches 56; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 2160 tattatgccacacttggtgtttgtgtgagataaacctttccagactcccaaaaaa 2219
      ||||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1001 ttctactaaaaatggtgagcaatgaagacaataattttctattttccnaaaaaa 1060

QY 2220 aaaaaaaaaaaaaaaaaaaaaa 2241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1061 aaaaaaaaaaaaaaaaaaaaaa 1082

RESULT 6
US-08-946-026-35
; Sequence 35, Application US/08946026
; Patent No. 6034218
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Micham, Jennifer L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,026
; FILING DATE: 07-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-946-026-35

Query Match          1.8%; Score 40.4; DB 3; Length 333;
Best Local Similarity 68.3%; Pred. No. 0.047;
Matches 56; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```

```
QY 2160 tattatgcccacatttggtgtttgtgtgagataaacctttccagactcccaaaaaa 2219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 TATGTTGGCAATCTCTCTGTTCTTAATGAATAAATGTTTTATATACCTTTAGAAAAA 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2220 aaaaaaiaaaaaaiaaaaaa 2241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 AAAAAAAAAAAAAAAAAAAAAA 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-328-111-757/c
; Sequence 757, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 757
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(578)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-757

Query Match 1.8%; Score 40.4; DB 4; Length 578;
Best Local Similarity 79.7%; Pred. No. 0.065;
Matches 47; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2183 tgtgtgagataaacctttccagactcccaaaaaaiaaaaaaiaaaaaa 2241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 TTTCTTAATCCCACTTTCTTACTCCCAAAAAAAAAAAAAAAAAAAAAA 10
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-09-227-357-83
; Sequence 83, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925

; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (664)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (716)
; OTHER INFORMATION: n equals a,t,g, or c
```

```

; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1319)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-83

```

Query Match	1.8%	Score 40.4;	DB 4;	Length 1977;
Best Local Similarity	75.8%;	Pred. No. 0.13;		
Matches 50;	Conservative 0;	Mismatches 16;	Indels 0;	Gaps
QY 2176	tgtgtttttgttgagataaaccttttcagactcccaaaaaa	aaaaa	aaaaa	aaaaa
Db 1876	tgtgtttggagacggaataaatgttttctcattcaaaaaa	aaaaa	aaaaa	aaaaa
QY 2236	aaaaaa	2241		
Db 1936	aaaaaa	1941		

```

RESULT      9
US-08-742-011-1
; Sequence 1, Application US/08742011
; Patent No. 5824504
; GENERAL INFORMATION:
; APPLICANT: Elshourbagy, Nabil A.
; APPLICANT: Bergsma, Derk J.
; APPLICANT: Ellis, Catherine E.
; TITLE OF INVENTION: Human 7-Transmembrane Receptor
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,011
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T.
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50020P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 332..1357
US-08-742-011-1

```

	Query Match	1.8%	Score 40.4	DB 1	Length 2338
	Best Local Similarity	68.3%	Pred. No. 0.14		
	Matches 56	Conservative 0	Mismatches 26	Indels 0	Gaps
QY	2160	tattatgccacacactttgtgtttgtgtgtgagataaaacctttccagactcccaaaaaa	2219		
Db	2147	TATTACTACCATATGAGTTTCTATAGCTAAGAAATAAAATCTTTAAAGTCTCAAAAAAAA	2206		

```

Qy  2220 aaaaaaaaaaaaaaaaaaaaaa 2241
    |oooooooooooooooooooo|
Db  2207 AAAAAAAAAAAAAAAAAAAAAA 2228

RESULT 10 .
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzpgt-F1s
US-08-232-463-14

```

	Query Match	1.88;	Score 40.4;	DB 1;	Length 7218;
	Best Local Similarity	6.1%;	Pred. NO. 0.28;		
	Matches 23;	Conservative 190;	Mismatches 161;	Indels 0;	Gaps 0;
Qy	1369	gggaagttagtagagccacatgtcctgcctgtgccaggtgagaaagcagacagcagg	1428
Db	1395	RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	1336
Qy	1429	agcactgtccccaagctggctgagcagatgactggtagagagtttgaaagtgccctgt	1488
Db	1335	RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	1276
Qy	1489	cgcagatctccacatcgagttcgaagtcgaaggggctgaaaattgagtagcgtgcaagct	1548
Db	1275	RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	1216
Qy	1549	qgtaaagggaatgcctgtgcaagcctcagccctgggaactctgcacqaggaagctggt	1608

Page 7

APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Shericca C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/449,315
 FILING DATE: 24-MAY-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/Pl/CSG 1727
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689

```

Query Match      1.8%; Score 40.2; DB 1; Length 1181;
Best Local Similarity 78.7%; Pred. No. 0.11;
Matches 48; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2181 ttgttgtagaataaacctttccagactcccaaaaaaaaaaaaaaaaaaaaaa 2240
Db 78 TGTATATAAGAGTAATCTTCCTTGCTTCTATAAAAAAAAAAAAAAAAAAAAA 19

QY 2241 a 2241
Db 18 A 18

RESULT 13
US-08-449-315-27/c
; Sequence 27, Application US/08449315
; Patent No. 5650505
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:15:13 ; Search time 5652.01 Seconds
(without alignments)
8578.153 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 2241
Sequence: 1 tatctctgtgtggggcccg.....aaaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues
Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2241	100.0	2241	US-09-743-237-1	Sequence 1, Appl
2	2066.4	92.2	2201	US-60-360-207-983	Sequence 983, App
3	1747	78.0	1861	US-09-743-237-2	Sequence 2, Appl
4	649.4	29.0	2638	US-09-643-893-287	Sequence 287, App
5	647.8	28.9	2134	US-09-743-237-3	Sequence 3, Appl
6	560.4	25.0	1770	US-09-471-275-5610	Sequence 5610, Ap
7	271.8	12.1	397	US-09-289-768-13239	Sequence 13239, A
8	271.8	12.1	397	US-09-939-397-13239	Sequence 13239, A
9	248.4	11.1	445	PCT-US01-01329-425	Sequence 425, App
10	248.4	11.1	445	PCT-US01-01339-1301	Sequence 1301, App
11	248.4	11.1	445	US-09-764-891-1301	Sequence 1301, Ap
12	248.4	11.1	445	US-10-080-090-425	Sequence 425, App
13	218	9.7	431	US-08-971-197-11928	Sequence 11928, A
14	218	9.7	431	US-08-971-197C-11928	Sequence 11928, A
15	175.4	7.8	4276	US-60-360-207-19806	Sequence 19806, A
16	168	7.5	3495	US-09-611-526-2894	Sequence 2894, Ap
17	157.4	7.0	3535	US-09-087-136-16	Sequence 16, Appl
18	157.4	7.0	3535	US-09-087-136-16	Sequence 16, Appl
19	157.4	7.0	3535	US-09-220-091-16	Sequence 16, Appl
20	152	6.8	352	US-09-515-128-15843	Sequence 15843, A
21	152	6.8	352	US-09-205-070-7224	Sequence 7224, Ap
22	152	6.8	352	US-09-340-623-7224	Sequence 7224, Ap
23	152	6.8	352	US-09-898-888-7224	Sequence 7224, Ap
24	152	6.8	352	US-09-898-888A-7224	Sequence 7224, Ap
25	152	6.8	795	PCT-US01-08656-5196	Sequence 5196, Ap
26	150.4	6.7	352	US-09-471-275-7634	Sequence 7634, Ap
27	123.6	5.5	444	US-09-404-549-314	Sequence 314, App
28	123.6	5.5	444	US-09-404-549A-314	Sequence 314, App
29	120.4	5.4	250	PCT-US00-26524B-2645	Sequence 2645, App
30	109.4	4.9	249	US-09-539-806-21625	Sequence 21625, A
31	105.4	4.7	1093	PCT-US01-08631-17949	Sequence 17949, A

32 103.8 4.6 465 17 US-09-362-510-45961 Sequence 45961, A
33 103.8 4.6 465 17 US-09-362-510A-45961 Sequence 45961, A
34 103.8 4.6 465 34 US-09-904-013-45961 Sequence 45961, A
35 92.8 4.1 2984 55 US-60-167-217-12947 Sequence 12947, A
36 92.8 4.1 3164 23 US-09-614-150-12896 Sequence 12896, A
37 92.8 4.1 3164 58 US-60-151-637-12934 Sequence 12934, A
38 92.4 4.1 302 14 US-09-013-812-1497 Sequence 1497, Ap
39 92.4 4.1 302 21 US-09-540-764-42635 Sequence 42635, A
40 92.4 4.1 302 42 US-60-036-570-1497 Sequence 1497, Ap
41 78.2 3.5 1716 28 US-09-708-427-33364 Sequence 33364, A
42 78.2 3.5 1812 28 US-09-708-427-27293 Sequence 27293, A
43 76.6 3.4 124 25 US-09-643-893-39 Sequence 39, App1
44 74.8 3.3 97814 20 US-09-534-859-253 Sequence 253, App
45 74.8 3.3 97814 31 US-09-803-736-253 Sequence 253, App

ALIGNMENTS

RESULT 1
US-09-743-237-1
; Sequence 1, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YOUNI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (651)..(1535)
US-09-743-237-1

Query Match 100.0%; Score 2241; DB 29; Length 2241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tatectgtggttgcccccgggagcaggctctcagcaggctcaggcaccacacagggata 60
Db 1 tatectgtggttgcccccgggagcaggctctcagcaggctcaggcaccacacagggata 60
Qy 61 cacagtgtgttctgctgctggtgactgtgactccaccaccctccgcccagcagg 120
Db 61 cacagtgtgttctgctgctggtgactgtgactccaccaccctccgcccagcagg 120
Qy 121 ctaggatagaacccaggccttttgcgtgtctgcagatagctctcagcctgtagtt 180
Db 121 ctaggatagaacccaggccttttgcgtgtctgcagatagctctcagcctgtagtt 180
Qy 181 ggggttggtgggagatttttttttttcacacaaagacttccattattgagatttt 240
Db 181 ggggttggtgggagatttttttttttcacacaaagacttccattattgagatttt 240
Qy 241 tcaattgatgatccccctctgtaagataaaggacagttcttttaaacctatgtagt 300
Db 241 tcaattgatgatccccctctgtaagataaaggacagttcttttaaacctatgtagt 300
Qy 301 ttatgaattctgtcttcaacattatgtaagctatatagcaattcttgaattgc 360
Db 301 ttatgaattctgtcttcaacattatgtaagctatatagcaattcttgaattgc 360

Qy 361 tatataacttagagaaacctctgattctctctcctctacatcctgagtgctaggtgtaca 420
Db 361 tatataacttagagaaacctctgattctctctcctctacatcctgagtgctaggtgtaca 420
Qy 421 ggggaaatcattttggtgagactccgatgaactactccaggttcccaagcagcaagc 480
Db 421 ggggaaatcattttggtgagactccgatgaactactccaggttcccaagcagcaagc 480
Qy 481 aagcaagaaaaagtgttgaatcaaaagaagcagggtggttagtgccagcgccagccctg 540
Db 481 aagcaagaaaaagtgttgaatcaaaagaagcagggtggttagtgccagcgccagccctg 540
Qy 541 aagacgcagctttccaggccctctggtcaggaatctctgttgcaggttcccatatccc 600
Db 541 aagacgcagctttccaggccctctggtcaggaatctctgttgcaggttcccatatccc 600
Qy 601 aggaggcagagaggcctccagctccctcgaaagaagactccagcccatggtgattt 660
Db 601 aggaggcagagaggcctccagctccctcgaaagaagactccagcccatggtgattt 660
Qy 661 gtcagctgaagaggcgcgcagatgctctgcatagaacaactgtggtgagggagctca 720
Db 661 gtcagctgaagaggcgcgcagatgctctgcatagaacaactgtggtgagggagctca 720
Qy 721 aagcgtccatctgttccatcagatcagatgacacagagcaggtttccctcagtcagc 780
Db 721 aagcgtccatctgttccatcagatcagatgacacagagcaggtttccctcagtcagc 780
Qy 781 ctaagccaatgacaacttttagtgggaagactctgcagtcacagcgaagttaaatctca 840
Db 781 ctaagccaatgacaacttttagtgggaagactctgcagtcacagcgaagttaaatctca 840
Qy 841 tcacacaggttgataatggagctctcccatcagctgcaatggggtgcttccctctg 900
Db 841 tcacacaggttgataatggagctctcccatcagctgcaatggggtgcttccctctg 900
Qy 901 gacctgtctgcaaggccaccacaaataaactctgtctgggtactgtgactgtcttcca 960
Db 901 gacctgtctgcaaggccaccacaaataaactctgtctgggtactgtgactgtcttcca 960
Qy 961 gggggactctgcacacagctcagctgcaacacacctgcgcacagctgagcgcgtcca 1020
Db 961 gggggactctgcacacagctcagctgcaacacacctgcgcacagctgagcgcgtcca 1020
Qy 1021 aagccataaaggcgtgtctgtatagaaatctctgaaagctttccaacacaaatgggaaag 1080
Db 1021 aagccataaaggcgtgtctgtatagaaatctctgaaagctttccaacacaaatgggaaag 1080
Qy 1081 gccgtctggagcgtctaaacttcgacacagcaaaagggtgcaactgaagcgtcaggct 1140
Db 1081 gccgtctggagcgtctaaacttcgacacagcaaaagggtgcaactgaagcgtcaggct 1140
Qy 1141 gccctgaagaactactgtgagtgctatgaggcaaaatcatgtttctccatttgcata 1200
Db 1141 gccctgaagaactactgtgagtgctatgaggcaaaatcatgtttctccatttgcata 1200
Qy 1201 gcattgtctgcaaaactatgaagaagtccagaaacaaaatgctgtatgagcacacccc 1260
Db 1201 gcattgtctgcaaaactatgaagaagtccagaaacaaaatgctgtatgagcacacccc 1260
Qy 1261 actaatgagcctggggactttgagagcagccattattgtcccccagccaaattctcag 1320
Db 1261 actaatgagcctggggactttgagagcagccattattgtcccccagccaaattctcag 1320
Qy 1321 gacctccaaactgagaaaaatagcagcgccttctcctgtatctcctgggagtagtg 1380
Db 1321 gacctccaaactgagaaaaatagcagcgccttctcctgtatctcctgggagtagtg 1380
Qy 1381 agccacatgtgctgctgctggccaggtgaggaagcagcagcagcagcagcagcagc 1440
Db 1381 agccacatgtgctgctgctggccaggtgaggaagcagcagcagcagcagcagcagc 1440
Qy 1441 caagcttgctgagcagatgatcctggagaggttttgaaggtgctgtgcagattctcc 1500

```

1441 caagcttgctgagcagatgatctctgagggagtttgaaagtgctgctgcagattctcc 1500
1501 acatcgagttcaagtcceagggtgtaaaattgagtagcgtgcagctggttaaaaggggaa 1560
1502 acatcgagttcaagtcceagggtgtaaaattgagtagcgtgcagctggttaaaaggggaa 1560
1503 tgcctgtggcaagctcagcctcagcctgggaatctgcacccaggaagctggtgcaccagggagga 1620
1504 tgcctgtggcaagctcagcctcagcctgggaatctgcacccaggaagctggtgcaccagggagga 1620
1505 gcaagggccgcgcacatcagcctcagcctgggaatctgcacccaggaagctggtgcaccagggagga 1680
1506 gcaagggccgcgcacatcagcctcagcctgggaatctgcacccaggaagctggtgcaccagggagga 1680
1507 gccagctactcaaggtatctcaagtcgaagcagcagcagcagcagcagcagcagcagcagcagcagc 1740
1508 gccagctactcaaggtatctcaagtcgaagcagcagcagcagcagcagcagcagcagcagcagcagc 1740
1509 tggccctcctgctccctggggagcctctggggagcctcctggggagcctcctggggagcctcctggggag 1800
1510 tggccctcctgctccctggggagcctctggggagcctcctggggagcctcctggggagcctcctggggag 1800
1511 gattttctactgtgttatgtgttggcttcaaatgtcttagtagtagtagtagtagtagtagtagtagtag 1860
1512 gattttctactgtgttatgtgttggcttcaaatgtcttagtagtagtagtagtagtagtagtagtagtag 1860
1513 ttattatgagcagcctcaagttagagagctagcctctctctctctctctctctctctctctctctctct 1920
1514 ttattatgagcagcctcaagttagagagctagcctctctctctctctctctctctctctctctctctct 1920
1515 atcacatacaagtcaggtggccatcaggggttttccagggcagcctctgacagggagat 1980
1516 atcacatacaagtcaggtggccatcaggggttttccagggcagcctctgacagggagat 1980
1517 atggaggagggttcgggttagagctgggtttgttggatttttggatttttggatttttggatttttggatt 2040
1518 atggaggagggttcgggttagagctgggtttgttggatttttggatttttggatttttggatttttggatt 2040
1519 gtattttctgctgaagtgagaaactgtctctctctctctctctctctctctctctctctctctctctct 2100
1520 gtattttctgctgaagtgagaaactgtctctctctctctctctctctctctctctctctctctctctct 2100
1521 ctgcacggtcgcctgctgaccagtcacagtcacagtcacagtcacagtcacagtcacagtcacagtcacag 2160
1522 ctgcacggtcgcctgctgaccagtcacagtcacagtcacagtcacagtcacagtcacagtcacagtcacag 2160
1523 attatgccacactttgtgtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2220
1524 attatgccacactttgtgtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2220
1525 aaaaaaaaaaaaaaaaaaaaaa 2241
1526 aaaaaaaaaaaaaaaaaaaaaa 2241
```

RESULT 2

```

US-60-360-207-983
; Sequence 983, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60360, 207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 983
; LENGTH: 2201
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-983
```

```

Query Match          92.2%; Score 2066.4; DB 75; Length 2201;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 2170; Conservative 0; Mismatches 16; Indels 14; Gaps 8;

Qy 20 gggcagcaggctctcagcaggtctcagcaggtccagccacacagggatcacacagtggtgtgttctctggc 79
Db 9 gggctgccccggcagcaggtctcagcaggtccagccacacagggatcacacagtggtgtgttctctggc 67
Qy 80 ctgttgacttctgactccaccaccaccctcccccagcagcagcagcagcagcagcagcagcagcagcagc 139
Db 68 ctgttgacttctgactccaccaccaccctcccccagcagcagcagcagcagcagcagcagcagcagcagc 127
Qy 140 ccttttgcgttcttcagcagtagtcttcagcctgttagtttgggttggctggcggagattt 199
Db 128 ccttttgcgttcttcagcagtagtcttcagcctgttagtttgggttggctggcggagattt 187
Qy 200 ttttttttcaacccaagacttccattatttggagatttttccagtttccagtttgatgatctcccc 259
Db 188 ttttttttcaacccaagacttccattatttggagatttttccagtttccagtttgatgatct-cccc 246
Qy 260 ctctgtaagataaaggacagcttctttaaaccctatgtagagtttttgatgaaattctgcttct 319
Db 247 ctctgtaaga caaggacagcttctttaaaccctatgtagagtttttgatgaaattctgcttct 306
Qy 320 caacatattgtctaaagctatagcaattcttgaatttgcataataacttaggagaacc 379
Db 307 caacatattgtctaaagctatagcaattcttgaatttgcataataacttaggagaacc 366
Qy 380 tctgattctcctcctctacatctctgagtgctaggtgtacaggggggaaatcatttttgggtg 439
Db 367 tctgattctcctcctctacatctctgagtgctaggtgtacaggggggaaatcatttttgggtg 426
Qy 440 agactccagataaactactgcccaggttcccagggcagcagcagcagcagcagcagcagcagcagcagcagc 499
Db 427 agactccagataaactactgcccaggttcccagggcagcagcagcagcagcagcagcagcagcagcagcagc 482
Qy 500 aatcaaaagacagcaggtgtagtgccagggcggcagcctgaaagcagcagcagcagcagcagcagcagcagc 559
Db 483 aatcaaaagacagcaggtgtagtgccagggcggcagcctgaaagcagcagcagcagcagcagcagcagcagc 542
Qy 560 cctctggtcaggaatctctgttcaagtttcccattatcccagagcagcagcagcagcagcagcagcagcagcagc 619
Db 543 cctctggtcaggaatctctgttcaagtttcccattatcccagagcagcagcagcagcagcagcagcagcagcagc 602
Qy 620 cagctccctcggaaagaaagactccagcccatggtgatttgcagctcagcagcagcagcagcagcagcagcagcagc 679
Db 603 cagctccctcggaaagaaagactccagcccatggtgatttgcagctcagcagcagcagcagcagcagcagcagcagc 662
Qy 680 ccagatgctctgcataagaaactgtgcccagggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 739
Db 663 ccagatgctctgcataagaaactgtgcccagggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 722
Qy 740 tcagtagcagtcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 799
Db 723 tcagtagcagtcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 782
Qy 800 agtgggaagacttctgcagtcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 859
Db 783 agtgggaagacttctgcagtcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 842
Qy 860 agctctcccatcagcagtcacatgggggcttccctctgagcagcagcagcagcagcagcagcagcagcagcagc 919
Db 843 agctctcccatcagcagtcacatgggggcttccctctgagcagcagcagcagcagcagcagcagcagcagcagc 902
Qy 920 acccaaaataactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 979
Db 903 acccaaaataactctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 962
Qy 980 ctgcagctgcacaaactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1039
Db 963 ctgcagctgcacaaactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1022
Qy 1040 tgatagaaatcctgaagcttttccacacacacacacacacacacacacacacacacacacacacacacacacacac 1099
```

```
Db 1023 tgaagaaatcctgaagctttcaaccacaaatgggaaaggcgtctgaggctgtaa 1082
Qy 1100 acttcgacacagaaagggtgcaactgtaagcgtcaagcgtcctgaagaactactgtga 1159
Db 1083 acttcgacacagaaagggtgcaactgtaagcgtcctgaagaactactgtga 1142
Qy 1160 gtgctatgag----gcaaaatcatgtgttcttccattgccaatgcatgcttgcacaaa 1215
Db 1143 gtgctatgaggcaagccaaatcatgtgttcttccattgccaatgcatgcttgcacaaa 1202
Qy 1216 actatgaagaagtccagaaacgaaatgctgatgagcaccaccccaactacatgagcctg 1275
Db 1203 actatgaagaagtccagaaacgaaatgctgatgagcaccaccccaactacatgagcctg 1262
Qy 1276 ggaactttgagacagaccattattgtcccaaggccaagtctcaggacctccaaaactga 1335
Db 1263 ggaactttgagacagaccattattgtcccaaggccaagtctcaggacctccaaaactga 1322
Qy 1336 gaaaaatagcgaggccttctctgtatctcctgggaagttagtgaggccacatgtgcct 1395
Db 1323 gaaaaatagcgaggccttctctgtatctcctgggaagttagtgaggccacatgtgcct 1382
Qy 1396 gctgtgcccagggtgaggaagcagagcagagcaactgttcccgaagcttggtgagc 1455
Db 1383 gctgtgcccagggtgaggaagcagagcagagcaactgttcccgaagcttggtgagc 1442
Qy 1456 agatgacctggaggagtttgggaaggtgctgtcgagattctccacatcgagtccaagt 1515
Db 1443 agatgacctggaggagtttgggaaggtgctgtcgagattctccacatcgagtccaagt 1502
Qy 1516 ccaaggggctgaaatagtagctgtgcaagctggtaaaagggaatgcctgtggcaagcc 1575
Db 1503 ccaaggggctgaaatagtagctgtgcaagctggtaaaagggaatgcctgtggcaagcc 1562
Qy 1576 tcagccctggaatctgcaccgggaagctggtgccaggaggagcagagggccgcgcacat 1635
Db 1563 tcagccctggaatctgcaccgggaagctggtgccaggaggagcagagggccgcgcacat 1622
Qy 1636 catggccaggtcagctgtgaggtcgtgagtgatctgcatggtactggccagcctactcaag 1695
Db 1623 catggccaggtcagctgtgaggtcgtgagtgatctgcatggtactggccagcctactcaag 1682
Qy 1696 gtatcctaaagtcaagcagcagagccaccctgggagtggaactggacactggccctcctgcc 1755
Db 1683 gtatcctaaagtcaagcagcagagccaccctgggagtggaactggacactggccctcctgcc 1742
Qy 1756 tggggaaggccctctggggaactcctgccctgcataaaagagggtgattttctacttgtt 1815
Db 1743 tggggaaggccctctggggaactcctgccctgcataaaagagggtgattttctacttgtt 1802
Qy 1816 gttatgtgttcttcaaatgcttagtagtacctccattcaagtattattatgagccagc 1875
Db 1803 gttatgtgttcttcaaatgcttagtagtacctccattcaagtattattatgagccagc 1862
Qy 1876 ctcaagttagagactaggtccttctcagggtggaactctgccaaatcacatacaagtcga 1935
Db 1863 ctcaagttagagactaggtccttctcagggtggaactctgccaaatcacatacaagtcga 1922
Qy 1936 ggtggccatcaggggtttttccaggccagccctgtgacagagagatatggga-ggggggtc 1994
Db 1923 ggtggccatcagggg-ctctccaggccagccctgtgacagagagatatggga-ggggggtc 1981
Qy 1995 ggggttagagctgggtttgttgatttttggcgtttt-cttctccgtgatttctgcttg 2053
Db 1982 ggggttagagctgggtttgttgatttttggcgtttt-cttctccgtgatttctgcttg 2041
Qy 2054 aagtgaagaaactgtctcctgccaa-cttttctccataaattactgtgcacggtcgc 2112
Db 2042 aagtgaagaaactgtctcctgccaaactcctgctccataaattactgtgcacggtcgc 2101
Qy 2113 ctgctgaccagtcacagtgaccctcagacacacagaagtgagggtggtttattatgccaca 2172
```

```
Db 2102 ctgctgaccagtcacagtgacctcagacaccagaagtgaggtyggtattatgccaca 2161
Qy 2173 cttgtgttttgtgtgagaataaaacctttccagactccc 2212
Db 2162 cttgtgttttgtgtgagaataaaacctttccagactccc 2201

RESULT 3
US-09-743-237-2
; Sequence 2, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YOSHI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (271)..(1155)
US-09-743-237-2
```

```
Query Match 78.0%; Score 1747; DB 29; Length 1861;
Best Local Similarity 100.0%; Pred. No. 7e-267;
Matches 1747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 495 gttgaatacaagaagcagggtgtagtgtagtgccaggcggcagccctgaagcagcagcttc 554
Db 115 gttgaatacaagaagcagggtgtagtgtagtgccaggcggcagccctgaagcagcagcttc 174
Qy 555 caggccctctggctcaggaaatcctgttccaagttcccatcatccacaggagcagagag 614
Db 175 caggccctctggctcaggaaatcctgttccaagttcccatcatccacaggagcagagag 234
Qy 615 gccctcagctgccctcggaagaagactccagcccatggtagttgtcagctgaaagga 674
Db 235 gccctcagctgccctcggaagaagactccagcccatggtagttgtcagctgaaagga 294
Qy 675 ggcgcacagatgctctgcatagacaactgtggcgcaggaggagctcaaaagcgtccatctg 734
Db 295 ggcgcacagatgctctgcatagacaactgtggcgcaggaggagctcaaaagcgtccatctg 354
Qy 735 cttccctcagtagatgaccagagcagcttccctcagtcagagctccctcaagccaatgaca 794
Db 355 cttccctcagtagatgaccagagcagcttccctcagtcagagctccctcaagccaatgaca 414
Qy 795 actttagtgggaagactcttccagtaaccaggaagttaaatctcatcacacaggttgat 854
Db 415 actttagtgggaagactcttccagtaaccaggaagttaaatctcatcacacaggttgat 474
Qy 855 aatggagctctcccatcagctgtcaatggggtgcctttccctctctggaactgctctgcaa 914
Db 475 aatggagctctcccatcagctgtcaatggggtgcctttccctctctggaactgctctgcaa 534
Qy 915 gggccaccccaaaaactctgtctgggtactgtgactgtcttccagcgggactctgc 974
Db 535 gggccaccccaaaaactctgtctgggtactgtgactgtcttccagcgggactctgc 594
Qy 975 aacagctgcagctgcacaacctgcgccatgagctgcagcgtcttcaaaagccaataagggcg 1034
Db 595 aacagctgcagctgcacaacctgcgccatgagctgcagcgtcttcaaaagccaataagggcg 654
```

QY 1035 tgccttgatagaatcctggaagctttccaaacccaaaggggaaagccgctgagact 1094
Db 655 tgccttgatagaatcctggaagctttccaaacccaaaggggaaagccgctgagact 714
QY 1095 gctaaacttcgacacagcaaaaggggtgaactgaagcgcagctgagcaactac 1154
Db 715 gctaaacttcgacacagcaaaaggggtgaactgaagcgcagctgagcaactac 774
QY 1155 tgccttgatagaatcctggaagctttccaaacccaaaggggaaagccgctgagact 1214
Db 775 tgccttgatagaatcctggaagctttccaaacccaaaggggaaagccgctgagact 834
QY 1215 aactatgaagaagtcacagcaaaaggggtgaactgaagcgcagctgagcaactac 1274
Db 835 aactatgaagaagtcacagcaaaaggggtgaactgaagcgcagctgagcaactac 894
QY 1275 ggggactttgagagcagcattattgttcccccagcagcagcttcccccagcagcagc 1334
Db 895 ggggactttgagagcagcattattgttcccccagcagcagcagcttcccccagcagcagc 954
QY 1335 agaaaaataagcagcagccttccctgtatctctggaagtagtgagcagcagcagcagc 1394
Db 955 agaaaaataagcagcagccttccctgtatctctggaagtagtgagcagcagcagcagc 1014
QY 1395 tgcctgctggccaggggtgaggaagcagcagcagcagcagcagcagcagcagcagcagc 1454
Db 1015 tgcctgctggccaggggtgaggaagcagcagcagcagcagcagcagcagcagcagcagc 1074
QY 1455 cagatgactcctggaaggttgggaaggtgctgctgcagcagcagcagcagcagcagcagc 1514
Db 1075 cagatgactcctggaaggttgggaaggtgctgctgcagcagcagcagcagcagcagcagc 1134
QY 1515 tccaggggctgaaatgagtagcagcagcagcagcagcagcagcagcagcagcagcagc 1574
Db 1135 tccaggggctgaaatgagtagcagcagcagcagcagcagcagcagcagcagcagcagc 1194
QY 1575 ctccagcctgggaatctgcacagcagcagcagcagcagcagcagcagcagcagcagcagc 1634
Db 1195 ctccagcctgggaatctgcacagcagcagcagcagcagcagcagcagcagcagcagcagc 1254
QY 1635 tcatggcaggtcagctgaggttgcagcagcagcagcagcagcagcagcagcagcagcagc 1694
Db 1255 tcatggcaggtcagctgaggttgcagcagcagcagcagcagcagcagcagcagcagcagc 1314
QY 1695 ggtatctaaagtcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1754
Db 1315 ggtatctaaagtcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1374
QY 1755 ctggggaggccctctgggagcctccctgcacaaagaggggtgatttctactgt 1814
Db 1375 ctggggaggccctctgggagcctccctgcacaaagaggggtgatttctactgt 1434
QY 1815 tcttatgttctgcttcaaatgcttagtagtagtagtagtagtagtagtagtagtagtagtag 1874
Db 1435 tcttatgttctgcttcaaatgcttagtagtagtagtagtagtagtagtagtagtagtagtag 1494
QY 1875 cctcaagttagagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1934
Db 1495 cctcaagttagagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1554
QY 1935 aggtggccatcaggggttttccagcagcagcagcagcagcagcagcagcagcagcagcagc 1994
Db 1555 aggtggccatcaggggttttccagcagcagcagcagcagcagcagcagcagcagcagcagc 1614
QY 1995 ggggttagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2054
Db 1615 ggggttagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1674
QY 2055 agtgagaaactgtctcctgcccacactttctccataatctacgctgagcagcagcagc 2114
Db 1675 agtgagaaactgtctcctgcccacactttctccataatctacgctgagcagcagcagc 1734
QY 2115 gctgaccagtcagctgacccctcagacacacagaggtgaggtgaggtgaggtgaggtgaggt 2174

Db 1735 gctgaccagtcagctgacccctcagacacagaggtgaggtgaggtgaggtgaggtgaggt 1794
QY 2175 tctgtttttgtgtgagataaaaccccttccagactcccaaaaaaaacacacacacacacac 2234
Db 1795 tctgtttttgtgtgagataaaaccccttccagactcccaaaaaaaacacacacacacacac 1854
QY 2235 aaaaaa 2241
Db 1855 aaaaaa 1861
RESULT 4
US-09-643-893-287
; Sequence 287, Application US/09643893
; GENERAL INFORMATION:
; APPLICANT: Silos-Santiago, Inmaculada
; APPLICANT: Dictefano, Peter
; APPLICANT: Curtis, Rory A.J.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES.
; FILE REFERENCE: 1600.1161-001
; CURRENT APPLICATION NUMBER: US/09/643,893
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/150,147
; NUMBER OF SEQ ID NOS: 412
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 287
; LENGTH: 2638
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2638)
; OTHER INFORMATION: n = A,T,C or G
US-09-643-893-287
Query Match 29.0%; Score 649.4; DB 25; Length 2638;
Best Local Similarity 76.2%; Pred. No. 3.9e-93;
Matches 830; Conservative 0; Mismatches 246; Indels 13; Gaps 2;
QY 495 gttgaaatacaag-aagcaggtggtgagtgccagcagcagcagcagcagcagcagcagcagc 553
Db 505 gttgaaatacaag-aagcaggtggtgagtgccagcagcagcagcagcagcagcagcagcagc 564
QY 554 ccaggccctctggtcaggaatcctgttgcaagttcccatcctcagcagcagcagcagcagcagc 613
Db 565 gcagaatctctgtcaggaatcctgttgcaagttcccatcctcagcagcagcagcagcagcagc 624
QY 614 ggcctcagcagcctcgcaggaagaaagcagcagcagcagcagcagcagcagcagcagcagcagc 673
Db 625 tgccctgctgcttcttcaagaagattcccaacccaatggtgagtagcagcagcagcagcagcagc 684
QY 674 aggcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 733
Db 685 gggcacacaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 744
QY 734 gcttcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 793
Db 745 ggttcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 804
QY 794 aactttagtggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 853
Db 805 tgccttagtgagggagatttttccagcagcagcagcagcagcagcagcagcagcagcagcagc 864
QY 854 taatggagcctccctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 913
Db 865 tgaggagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 924
QY 914 agggccaccccaaaataactctgtctgggtgaggtgaggtgaggtgaggtgaggtgaggtgaggt 973

Db 1270 cactgagtttaactaaggatgaaatgagtagagtataaagtgatgcatgtt 1329
Qy 1562 gctgtgtgc 1570
Db 1330 gattttgtc 1338

RESULT 6

US-09-471-275-5610/C
; Sequence 5610, Application US/09471275
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Contigs Obtained
; FILE REFERENCE: 782
; CURRENT APPLICATION NUMBER: US/09/471,275
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/235,076
; EARLIER FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: US 09/234,611
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: US 09/240,371
; EARLIER FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: US 09/277,227
; EARLIER FILING DATE: 1999-03-25
; EARLIER APPLICATION NUMBER: US 09/271,490
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: US 09/293,972
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: US 09/274,861
; EARLIER FILING DATE: 1999-03-23
; EARLIER APPLICATION NUMBER: US 60/125,453
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: US 60/126,605
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: US 09/306,350
; EARLIER FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 09/399,720
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER: US 09/404,284
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER:
; EARLIER FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 10451
; SOFTWARE: PLCT_genes Version 1.0
; SEQ ID NO 5610
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1771)...(874)
; OTHER INFORMATION: similar to g14581563 in the genepept database release 114,
; OTHER INFORMATION: Run with FASTX 3.3t00, default parameters
US-09-471-275-5610

Query Match 25.0%; Score 560.4; DB 18; Length 1770;
Best Local Similarity 76.1%; Pred. No. 4.7e-79;
Matches 709; Conservative 0; Mismatches 211; Indels 12; Gaps 1;
Qy 651 atgtgtattgtcagctgaaaggagggccagatgctctcatagacaaactgtggcg 710
Db 1770 ATGTGTATGCCAATGAAAGGGGGCACAAATGCTATATAGACAATTCAGAAC 1711
Qy 711 agggagctcaagcgctccatctctctcagtacgaccagagcagttccctcag 770
Db 1710 AGACAATAAAGACCTCCATTTGGTCTCAGTATCAAGATCAAAATAATATCTAC 1651
Qy 771 tcagagctccctaaagccaatgacaactttaggggaagactctgccagtcaggaag 830
Db 1650 TCAGATGTCCTAAACCAATCACTGCTTTAGTAGGAGATTTTCCAGCATCAACAAA 1591

Qy 831 ttaactctcatcacacaggttgataaaggagctctcccatcagctgtcctcaatgggctgccc 890
Db 1590 TTAANTCTCATTTACACAACAACCTTGGAGGAGCCCTTACCATTGGTAGTCAACGGGTCTGCT 1531
Qy 891 ttccctctgacactgctctgcaaggccaccaccccaaaataaactctgtctgggtactgtgac 950
Db 1530 TTCCCTCGGGATCAACTCTTCCAGGACCACCAAAAATAAATTTGGCTGGGTACTGTGAC 1471
Qy 951 tgcttctccagcgggactctctgcaacagctgca-----gctgcaacacactg 998
Db 1470 TGCTTTGCCAGTGGGACTTTTGGCAACAACATGCAATTTAATAATTTGTGCAACAACCTTG 1411
Qy 999 cgccatgagctcgagcgtctcaaaagccataaaggcgtctctgtatagataaactctgaaact 1058
Db 1410 CATCATGATATTGAACGGTTTAAAGCCATTAAAGCCATTGCTTTGGTAGAAAATCCAGAACT 1351
Qy 1059 ttccaacccaaaatgggaaaggccgtctgagagctgtctaaacttcgacacagcaaggg 1118
Db 1350 TTCCAGCCAAAATTTGGGAAGGCCAATTTGGCAATGTCAAGCCCCAGCACACAACAGGG 1291
Qy 1119 tgcaactgtgaagcgtcagcgtgctgaagaactactgtgagtgctatgagggccaaatc 1178
Db 1290 TGCAACTGCAGGAGGTGAGGCTGCTGAAGAATTTACTGCGAGTGTCTATGAGGCCCAAAAT 1231
Qy 1179 atgttctctccatttgcataatgcttgcataaactgtgcaaaactatgaagaagtcagaaacga 1238
Db 1230 ATGTGCTCTTCTATTGCAAAATGCTGTTGCAAAAATTTATGAAGAAGCCCAAGCA 1171
Qy 1239 aaaatgctgatgagcacacccactacatgagcctgggactttgagagcagccattat 1298
Db 1170 AAGACACTAATGAGCATGCGCAAACTTACATGAGACTGCGAGGTTTGGGAAGGCGCATTTAC 1111
Qy 1299 ttgtcccccagccaagttctcagggactccaaaactgagaaaaatagcagggcctctccc 1358
Db 1110 CTGCCACCACCAAGAAATTTTCAGGACTTCCAAAGATTCAGTACGATAGCGGCTTCTCTCA 1051
Qy 1359 tgtatctctgggaagtgtgagggccacatgtgctgcctgctggtgcccagggtgagaa 1418
Db 1050 TGCATCTCTCTGGAGGTGGTGGAGGCCACATGCGGCTGCCCTGCTGCTGCTCAGGAGAGAG 991
Qy 1419 gcagagcaggagcactgttcccccaagcttgctgagcagatgatcctggagagtttggaa 1478
Db 990 GCCGAGAAAGAACACTGCTCCCAAGTGCCTGGCAGAGCAGATGATCCTGGAGGAATTTGGA 931
Qy 1479 aggtgctgtcgcagattctccacatcagttcagttcagttcagttcagttcagttcagtt 1538
Db 930 AGGTGCTTTATCACAGATTTCTCCACACTGAGTTTAAATCTAAGGGATTGAAAATGAGTAG 871
Qy 1539 cgtgcaagctgttaaagggggaatgcctgtgac 1570
Db 870 AGTATAAGTGTGAATGATGCTGTTGATTTTGTG 839

RESULT 7

US-09-289-768-13239/C
; Sequence 13239, Application US/09289768
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-765
; CURRENT APPLICATION NUMBER: US/09/289,768
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 39996
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13239
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(397)

OTHER INFORMATION: n = A,T,C or G
US-09-289-768-13239

Query Match 12.1%; Score 271.8; DB 16; Length 397;
Best Local Similarity 80.5%; Pred. No. 2e-33; 77; Indels 0; Gaps 0;
Matches 318; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1119 tgcaactgtaagcgcctcagcgtcctgaagaactactgtgagtgctatgagcccaaatc 1178
DB 396 TGCAACTGCAGGAGGTCAGGCTGCCTGAAGATTACTGCGAGTGCTATGAGGCCCAAAAT 337

QY 1179 atgtgttcttcattgcaaatgcatgttgcataaaactatgaagaagtccagaacga 1238
DB 336 ATGTGTTCTTCTATTGCAAAATGCAATGTGTTGCAAAATATTATGAAGAAAGCCAGAACGA 277

QY 1239 aaaaactgatgacacacccactacatgagcctggagcttgagagcagcattat 1298
DB 276 AAGACACTAATGAGCATGCAAACTACATGCAGACTGGAGGTTTGAAGCGCAGCCATTAC 217

QY 1299 ttgtcccccagcaagtctcagagacctccaaactgagaaaaataggcgcctctcc 1358
DB 216 CTGCCACCAACGAATTTTCAGAGCTTCCAAGATTTCAGTCACGATAGCGGCTTCTCTCA 157

QY 1359 tgatatctctggaaagtagtgagggccacatgtgcctgctgctggcccccaggtgagaa 1418
DB 156 TGCACTCTCTGGAGGTGGTGGAGGCCACATGGCCTGCTTCTCAGGAGAGAGAG 97

QY 1419 gcagagcagagcactgttccccaaactggctgagcagacatgctgagagattgga 1478
DB 96 GCCGAGAAAGAACCCCTGCTCCCAAGTGCCTGGCAGACGATGATCTCTGGAGGAATTTGA 37

QY 1479 aggtgctgtcgcagattctccacatcgagttcaa 1513
DB 36 AGGTGCTTATCACAGATTCTCCACACTGAGTTTAA 2

RESULT 8

US-09-939-397-13239/c
Sequence 13239, Application US/09939397
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-765
CURRENT APPLICATION NUMBER: US/09/939,397
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: 09/289,768
PRIOR FILING DATE: 1999-04-08
NUMBER OF SEQ ID NOS: 39996
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 13239
LENGTH: 397
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(397)
OTHER INFORMATION: n = A,T,C or G
US-09-939-397-13239

Query Match 12.1%; Score 271.8; DB 35; Length 397;
Best Local Similarity 80.5%; Pred. No. 2e-33;
Matches 318; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1119 tgcaactgtaagcgcctcagcgtcctgaagaactactgtgagtgctatgagcccaaatc 1178
DB 396 TGCAACTGCAGGAGGTCAGGCTGCCTGAAGATTACTGCGAGTGCTATGAGGCCCAAAAT 337

QY 1179 atgtgttcttcattgcaaatgcatgttgcataaaactatgaagaagtccagaacga 1238
DB 336 ATGTGTTCTTCTATTGCAAAATGCAATGTGTTGCAAAATATTATGAAGAAAGCCAGAACGA 277

QY 1239 aaatgtgtgatgacacacccactacatgagcctggggaactttgagagcagcattat 1298
DB 276 AAGACACTAATGAGCATGCAAACTACATGCAGACTGGAGGTTTGGAGGCGCAGCCATTAC 217

QY 1299 ttgtcccccagcaagtctcagagacctccaaactgagaaaaataggcgcctctcc 1358
DB 216 CTGCCACCAACGAATTTTCAGAGCTTCCAAGATTTCAGTCACGATAGCGGCTTCTCTCA 157

QY 1359 tgatatctctggaaagtagtgagggccacatgtgcctgctgctggcccccaggtgagaa 1418
DB 156 TGCACTCTCTGGAGGTGGTGGAGGCCACATGGCCTGCTTCTCAGGAGAGAGAG 97

QY 1419 gcagagcagagcactgttccccaaactggctgagcagacatgctgagagattgga 1478
DB 96 GCCGAGAAAGAACCCCTGCTCCCAAGTGCCTGGCAGACGATGATCTCTGGAGGAATTTGA 37

QY 1479 aggtgctgtcgcagattctccacatcgagttcaa 1513
DB 36 AGGTGCTTATCACAGATTCTCCACACTGAGTTTAA 2

RESULT 9

PCT-US01-01329-425
Sequence 425, Application PC/TUS0101329
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc., et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA120PCT
CURRENT APPLICATION NUMBER: PCT/US01/01329
CURRENT FILING DATE: 2001-01-17
NUMBER OF SEQ ID NOS: 3506
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 425
LENGTH: 445
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (431)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
PCT-US01-01329-425

Query Match 11.1%; Score 248.4; DB 1; Length 445;
Best Local Similarity 73.9%; Pred. No. 1e-29;
Matches 312; Conservative 1; Mismatches 109; Indels 0; Gaps 0;

QY 495 gttgaaatcaagaagcaggtgtagtggtccaggcggcagccctgaaagcagcagcttc 554
DB 24 gttgaaatcaagaagcaggtgtagtggtccactactacaagtaataatccggaagaagcaacttg 83

QY 555 caggccctctggtcaggaatctctgttgcagctcccatcctccagagcagagagag 614
DB 84 cagaactcttctgtcaggaatctctgttgcagctcccatcctccagagagagagat 143

QY 615 gctccagctgcccctgggaagaagactccagcccatcctggtgattgtcagctgaaagga 674
DB 144 gctccgtgttctcttaagaagaagattcccaaccatgggtgatagccaatgaaaggg 203

QY 675 ggcgccagatgctctgcatagacaactgtggcggaggagagctcaaaagcgtccatctg 734
DB 204 ggcacacaatgctatgtagacaattctagacacagagaaactaaagcactccattg 263

QY 735 ctctcctcagtaoatgaccagagcaggttctccctcagtcagagctcctcaagcagaacga 794
DB 264 gttcctcagtatcaagatcaaaataatattctacagtcagatgtccctaaacaaatgact 323

QY 795 actttagtggaagactctcgcagtcacagcgaagttaaatctcatcacacaggttgat 854

Db 324 gctttagtaggagatttttccagcatcaacaaaattaaattctcattacacacaactt 383
Qy 855 aatggagcttccatcaagctgtcaatgggctgccttccctcttgagcctgctctgcaa 914
Db 384 gaggagccttaccatcggtagtcaacgggtctgtcttccctcggtgntcaactntcca 443
Qy 915 gg 916
Db 444 gg 445

RESULT 10

PCT-US01-01339-1301
: Sequence 1301, Application PC/TUS0101339

: GENERAL INFORMATION:

: APPLICANT: Human Genome Sciences, Inc., et al.

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

: FILE REFERENCE: PC006PCT

: CURRENT APPLICATION NUMBER: PCT/US01/01339

: CURRENT FILING DATE: 2001-03-17

: NUMBER OF SEQ ID NOS: 10231

: SOFTWARE: PatentIn Ver. 2.0

: SEQ ID NO 1301

: LENGTH: 445

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: NAME/KEY: SITE

: LOCATION: (431)

: OTHER INFORMATION: n equals a,t,g, or c

: NAME/KEY: SITE

: LOCATION: (438)

: OTHER INFORMATION: n equals a,t,g, or c

: PCT-US01-01339-1301

Query Match 11.1%; Score 248.4; DB 1; Length 445;
Best Local Similarity 73.9%; Pred. No. 1e-29;
Matches 312; Conservative 1; Mismatches 109; Indels 0; Gaps 0;

Qy 495 gttgaaatcaagaagcagggtgtgtgctccaggcggcagccctgaagacgcagcttcc 554
Db 24 gttgaaatcaagaagcagggtgtgtgctccaggcggcagccctgaagacgcagcttcc 83
Qy 555 caggccctctgctcaggaaatcctgttgcaagtcccatcctccaggaagcagagag 614
Db 84 cagaatctctgtcaggaaatcctgttgcaagtcccatcctccaggaagcagagag 143
Qy 615 gcttccagctgccctcggaagaagactccagcccatggtgattgtcagctgaaagga 674
Db 144 gcttccctgctgttctttaaagaagattcccaaccataatggtgatacgcaattgaaagg 203
Qy 675 gggccccaagctgctgtcagacaaactgtggcgagggagctcaaaagcgtccatctg 734
Db 204 ggcacacaaatgctatgtatagacaaatctagacraagaactaaaagcactccatttg 263
Qy 735 ctctcagtagcagatgaccagagcaggtttccctcagtcagagctcccttaagccaatgaca 794
Db 264 gttcctcagtagcagatgaccagagcaggtttccctcagtcagagctcccttaagccaatgact 323
Qy 795 actttagtggaagactctcgcagtagcagcaggaagttaaattctcattcacacaggtgat 854
Db 324 gctttagtaggagatttttccagcatcaacaaaattaaattctcattacacacaactt 383
Qy 855 aatggagcttcccatcagctgtcaatgggctgccttccctctggacctgctctgcaa 914
Db 384 gaggagccttaccatcggtagtcaacgggtctgttccctcggtgntcaactntcca 443
Qy 915 gg 916
Db 444 gg 445

RESULT 11

US-09-764-891-1301
: Sequence 1301, Application US/09764891

: GENERAL INFORMATION:

: APPLICANT: Rosen et al.

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

: FILE REFERENCE: PC006

: CURRENT APPLICATION NUMBER: US/09764,891

: CURRENT FILING DATE: 2001-01-17

: Prior application data removed - consult PALM or file wrapper

: NUMBER OF SEQ ID NOS: 10231

: SOFTWARE: PatentIn Ver. 2.0

: SEQ ID NO 1301

: LENGTH: 445

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: NAME/KEY: SITE

: LOCATION: (431)

: OTHER INFORMATION: n equals a,t,g, or c

: NAME/KEY: SITE

: LOCATION: (438)

: OTHER INFORMATION: n equals a,t,g, or c

: US-09-764-891-1301

Query Match 11.1%; Score 248.4; DB 30; Length 445;
Best Local Similarity 73.9%; Pred. No. 1e-29;
Matches 312; Conservative 1; Mismatches 109; Indels 0; Gaps 0;

Qy 495 gttgaaatcaagaagcagggtgtgtgctccaggcggcagccctgaagacgcagcttcc 554
Db 24 gttgaaatcaagaagcagggtgtgtgctccaggcggcagccctgaagacgcagcttcc 83
Qy 555 caggccctctgctcaggaaatcctgttgcaagtcccatcctccaggaagcagagag 614
Db 84 cagaatctctgtcaggaaatcctgttgcaagtcccatcctccaggaagcagagag 143
Qy 615 gcttccagctgccctcggaagaagactccagcccatggtgattgtcagctgaaagga 674
Db 144 gcttccctgctgttctttaaagaagattcccaaccataatggtgatacgcaattgaaagg 203
Qy 675 gggccccaagctgctgtcagacaaactgtggcgagggagctcaaaagcgtccatctg 734
Db 204 ggcacacaaatgctatgtatagacaaatctagacraagaactaaaagcactccatttg 263
Qy 735 ctctcagtagcagatgaccagagcaggtttccctcagtcagagctcccttaagccaatgaca 794
Db 264 gttcctcagtagcagatgaccagagcaggtttccctcagtcagagctcccttaagccaatgact 323
Qy 795 actttagtggaagactctcgcagtagcagcaggaagttaaattctcattcacacaggtgat 854
Db 324 gctttagtaggagatttttccagcatcaacaaaattaaattctcattacacacaactt 383
Qy 855 aatggagcttcccatcagctgtcaatgggctgccttccctctggacctgctctgcaa 914
Db 384 gaggagccttaccatcggtagtcaacgggtctgttccctcggtgntcaactntcca 443
Qy 915 gg 916
Db 444 gg 445

RESULT 12

US-10-080-090-425
: Sequence 425, Application US/10080090

: GENERAL INFORMATION:

: APPLICANT: Rosen et al.

: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

: FILE REFERENCE: PA120C1

: CURRENT APPLICATION NUMBER: US/10/080,090

: CURRENT FILING DATE: 2002-02-22

: Prior Application removed - See File Wrapper or Palm

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 09:13:13 ; Search time 690.24 Seconds
(without alignments)
5466.537 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 2241

Sequence: 1 tatctctgggtggcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1156321 seqs, 841861299 residues

Total number of hits satisfying chosen parameters: 2312642

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_New:

- 1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	647.8	28.9	2134	1	PCT-US02-10421-2790
2	647.8	28.9	2134	6	US-10-112-699-2790
3	647.8	28.9	2134	6	US-10-172-118-1104
4	279.4	12.5	486	1	PCT-US02-10421-2370
5	279.4	12.5	486	6	US-10-112-699-2370
6	152.6	6.8	352	5	US-09-721-544-15843
7	120.4	5.4	250	6	US-10-106-698-2655
8	72.8	3.2	1684	6	US-10-155-881-3628
9	66.4	3.0	1254	6	US-10-155-881-3943
10	52.8	2.4	763	6	US-10-155-881-3101
11	51.8	2.3	381	6	US-10-138-145-301
12	47.6	2.1	530	6	US-10-027-632-221356
13	47.4	2.1	249	5	US-09-975-254-20153
14	46	2.1	2022	1	PCT-US02-17315-3
15	46	2.1	2022	6	US-10-160-758-3
16	44	2.0	478	5	US-09-918-995-15466
17	43.8	2.0	625	6	US-10-137-321-97
18	43.4	1.9	2043	6	US-10-105-299-2661
19	43.4	1.9	2043	6	US-10-106-698-131
20	43.2	1.9	1098	6	US-10-162-758-42
21	43	1.9	939	6	US-10-119-480-77
22	43	1.9	1241	6	US-10-102-806-140
23	43	1.9	1241	6	US-10-144-850-144
24	42.8	1.9	476	5	US-09-918-995-15867
25	42.6	1.9	1162	6	US-10-106-698-764
26	42.4	1.9	1024	6	US-10-123-155-198

RESULT 1
PCT-US02-10421-2790
; Sequence 2790, Application PC/TUS0210421
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Siquing
; APPLICANT: Chaitanya S.
; APPLICANT: Gaiger, Alexander
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.565PC
; CURRENT APPLICATION NUMBER: PCT/US02/10421
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 2959
; SEQ ID NO 2790
; LENGTH: 2134
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US02-10421-2790

ALIGNMENTS

27	42.4	1.9	1024	6	US-10-137-871-198	Sequence 198, App
28	42.4	1.9	1024	6	US-10-141-761-198	Sequence 198, App
29	42.4	1.9	1024	6	US-10-140-864-198	Sequence 198, App
30	42.4	1.9	1024	6	US-10-140-923-198	Sequence 198, App
31	42.4	1.9	1024	6	US-10-141-756-198	Sequence 198, App
32	42.4	1.9	1024	6	US-10-141-759-198	Sequence 198, App
33	42.4	1.9	1024	6	US-10-140-472-198	Sequence 198, App
34	42.4	1.9	1024	6	US-10-140-805-198	Sequence 198, App
35	42.4	1.9	1024	6	US-10-142-885-198	Sequence 198, App
36	42.4	1.9	1024	6	US-10-146-731-198	Sequence 198, App
37	42.4	1.9	1024	6	US-10-142-426-198	Sequence 198, App
38	42.4	1.9	1024	6	US-10-158-790-198	Sequence 198, App
39	42.2	1.9	445	7	US-60-382-445-256	Sequence 256, App
40	42.2	1.9	831	6	US-10-172-118-1717	Sequence 1717, Ap
41	42	1.9	2387	6	US-10-105-299-3006	Sequence 3006, Ap
42	42	1.9	2591	6	US-10-137-757-64	Sequence 64, Appl
43	41.8	1.9	317	1	PCT-US02-12612-1260	Sequence 1260, Ap
44	41.8	1.9	317	6	US-10-125-968-1260	Sequence 1260, Ap
45	41.8	1.9	386	1	PCT-US02-12612-1243	Sequence 1243, Ap

Query Match 28.9%; Score 647.8; DB 1; Length 2134;
Best Local Similarity 76.1%; Pred. No. 9.4e-103;
Matches 829; Conservative 0; Mismatches 247; Indels 13; Gaps 2;

Qy	495	gttgaatacaag-aagcagggtgtagtgccaggcggcagccctgaagcagcagcttt	553
Db	250	gttgaatacaagtaagcagggtgtagtgccaggcggcagccctgaagcagcagcttt	309
Qy	554	ccaggccctctgctcaggaatcctgtgcaagttccatccatccaggaggcagagga	613
Db	310	gcagaatctctgctcaggaatcctgtgcaagttccatccatccaggaggcagagga	369
Qy	614	ggcctcagctgcctcggagaagaagactcagcccatcgtgtagttgtagcgaagg	673
Db	370	tgcctcgtctctcttaagaaagattccacccaatggtgatgtccaatgaaagg	429
Qy	674	agggccagatgctctgcatagacaactgtggcgcgaggaggtcgaagcgtccatct	733
Db	430	gggcacacaaatgtagatagacaatcttagacaagaagaactaaagcaccatt	489
Qy	734	gcttcctcagtagcagcagcagcttccctcctcagcagctccctaaagcacaatgac	793
Db	490	ggcttcctcagtagcagcagcagcttccctcctcagcagctccctaaagcacaatgac	549
Qy	794	aacttagtggaagactctgccagtagcagcaggaagtaaatctcatcacaggttga	853
Db	550	tgcttagtggaagactctgccagtagcagcaggaagtaaatctcatcacaggttga	609

Query Match 12.5%; Score 279.4; DB 1; Length 486;

Best Local Similarity 76.4%; Pred. No. 2.4e-39;

Matches 343; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 546 gcaagcttccagggccctctgctcaggaatcctgttgcaagttcccatccacagag 605
 DB 1 gcaacttgcagaatctctgtcaggaatcctgttgcaagttcccatccacagag 60
 QY 606 gcaagagagccctccagctgcccctggaagaagagactccagcccatggtgattgtcag 665
 DB 61 ctgagagagctcctcctgctgtctcttaagaagaagattccaaacccaatggtgatgccaa 120
 QY 666 ctgaaaggaggccagagctctctgctatagacaactgtggcgaggagctcgaagcg 725
 DB 121 ttgaaggaggccacacaaatgctatgataagacaattctagaacaagagaactaaagca 180
 QY 726 ctccatctctctcagtagcagcagagcaggttccctcagtcagagctccctaag 785
 DB 181 ctccattggttctcagtagcagcagagcaggttccctcagtcagagctccctaag 240
 QY 786 ccaatgacaactttagtggaagactctgcccagtagcagagcaggttccctcagtcagagctccctaag 845
 DB 241 ccaatgactgttttagtggaagactctgcccagtagcagagcaggttccctcagtcagagctccctaag 300
 QY 846 caggttgataatggagctctccatcagctgtcgaatggggtgctctccctcagagct 905
 DB 301 caacaacttgaggagccttaccatcggtagtcacacgggtctgtcttccctcagagct 360
 QY 906 gacttctgcaacagctgcagctgcaacaa 994
 DB 421 gacttttgcaacaactgcaattgtaataa 449

RESULT 5

US-10-112-699-2370

; Sequence 2370, Application US/10112699

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Wang, Siqing

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Gaiger, Alexander

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF CANCER

; FILE REFERENCE: 210121.565

; CURRENT APPLICATION NUMBER: US/10/112,699

; CURRENT FILING DATE: 2002-03-28

; NUMBER OF SEQ ID NOS: 2959

; SEQ ID NO 2370

; LENGTH: 486

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 484

; OTHER INFORMATION: n = A,T,C or G

US-10-112-699-2370

Query Match

Best Local Similarity 12.5%; Score 279.4; DB 6; Length 486;

Matches 343; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 546 gcaagcttccagggccctctgctcaggaatcctgttgcaagttcccatccacagag 605
 DB 1 gcaacttgcagaatctctgtcaggaatcctgttgcaagttcccatccacagag 60
 QY 606 gcaagagagccctccagctgcccctggaagaagagactccagcccatggtgattgtcag 665
 DB 61 ctgagagagctcctcctgctgtctcttaagaagaagattccaaacccaatggtgatgccaa 120

QY 666 ctgaaaggaggccagagctgtctgctatagacaactgtggcgagaggagctcaagcg 725
 DB 121 ttgaaaggaggccagacaatgctatgataagacaatcttagacaagagaactcaagca 180
 QY 726 ctccatctctcctcagtagcagcagagcaggttccctcagtcagagctccctaag 785
 DB 181 ctccattggttctcagtagcagcagagcaggttccctcagtcagagctccctaag 240
 QY 786 ccaatgacaactttagtggaagactctgcccagtagcagagcaggttccctcagtcagagctccctaag 845
 DB 241 ccaatgactgttttagtggaagactctgcccagtagcagagcaggttccctcagtcagagctccctaag 300
 QY 846 caggttgataatggagctctccatcagctgtcgaatggggtgctctccctcagagct 905
 DB 301 caacaacttgaggagccttaccatcggtagtcacacgggtctgtcttccctcagagct 360
 QY 906 gctctgcaaggccaccccaataactgtctgtgggtgactgtgactgtcttccctcagagcg 965
 DB 361 actcttccaggaccaccccaataactgttgggtgactgtgactgttccctcagagcg 420
 QY 966 gacttctgcaacagctgcagctgcaacaa 994
 DB 421 gacttttgcaacaactgcaattgtaataa 449

RESULT 6

US-09-721-544-15843

; Sequence 15843, Application US/09721544

; GENERAL INFORMATION:

; APPLICANT: Arterburn, Matthew

; APPLICANT: Asghari, Vida

; APPLICANT: Damavandi, Simin

; APPLICANT: Dickson, Mark

; APPLICANT: Drake, Jim

; APPLICANT: Drmanac, Radoje

; APPLICANT: Engleman, Carrie

; APPLICANT: Faulkner, Brandy

; APPLICANT: Garcia, Veronica

; APPLICANT: Giedt, Gretchen

; APPLICANT: Hunter, Kelly

; APPLICANT: Jessen, Aaron

; APPLICANT: Jones, Lee

; APPLICANT: Kita, David

; APPLICANT: Labat, Ivan

; APPLICANT: Laroya, Mimi

; APPLICANT: Lomelli, Michelle

; APPLICANT: Nguyen, Phuong

; APPLICANT: Nogra, Margie

; APPLICANT: Palencia, Servando

; APPLICANT: Raisi, Fariba

; APPLICANT: Smith, Benjamin

; APPLICANT: Tkach, Joe

; APPLICANT: Tran, Lien

; APPLICANT: Verna, Ron

; APPLICANT: Yang, Fei

; APPLICANT: Yim, Kenneth

; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained

; TITLE OF INVENTION: From a cDNA Library of Fetal Liver-Spleen

; FILE REFERENCE: 728CIP

; CURRENT APPLICATION NUMBER: US/09/721,544

; CURRENT FILING DATE: 2000-11-21

; PRIOR APPLICATION NUMBER: 09/515,128

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: 09/034,341

; PRIOR FILING DATE: 1998-02-13

; NUMBER OF SEQ ID NOS: 24489

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 15843

; LENGTH: 352

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:


```

; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3943
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Zea mays
; US-10-155-881-3943

Query Match          3.0%; Score 66.4; DB 6; Length 1254;
Best Local Similarity 68.9%; Pred. No. 0.011;
Matches 91; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1107 cacagcaagggtgcaactgtaagcgtcagctgctgaagaactcactgtgaagtcttat 1166
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 catacccccaggtgtcactgcaagaagtcaggtgtgtcttaagaagaaactgtgaattttt 80
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1167 gagcccaaaatcatgtgtcttccattgcataatgcattgttcgcaaaactatgaagaa 1226
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 caagcgaatatcttgcctccagaagattgtaaatgtatgtgattgcaaaactatgaagg 140
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1227 agtcagaaacga 1238
      ||| ||| |||
Db 141 agtgaggagcta 152
      ||| ||| |||

RESULT 10
US-10-155-881-3101
; Sequence 3101, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3101
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Zea mays
; US-10-155-881-3101

Query Match          2.4%; Score 52.8; DB 6; Length 763;
Best Local Similarity 59.2%; Pred. No. 2.3;
Matches 90; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1093 ctgctaaacttcagacagcaaggggtgcaactgtaagcgtcagcgtcgtgaagaact 1152
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 582 ctcttcaactccaccaagaagaggttgcaactgcaagaagtcgaagtgcgtcctaagaat 641
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1153 actgtgagtcctatgagcccaaaatcatgtgtcttccattgcataatgcattgcttga 1212
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 642 actgtgattgtaccaggtggtgagctggatgtctcttgttttgcctgctgacgtgccc 701
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1213 aaactatgaagaagcccaagcaagcaaaaatg 1244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 702 agaacccttttgggaaaaatgaaggcataatg 733
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-10-138-145-301
; Sequence 301, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: ShenK, Michael Andrew

```

```

; APPLICANT: Gibson, John Bryan
; APPLICANT: Norriss, Michael Geoffrey
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use.
; FILE REFERENCE: 11000.10580
; CURRENT APPLICATION NUMBER: US/10/138,145
; CURRENT FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 1652
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 301
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(381)
; OTHER INFORMATION: n = A,T,C or G
; US-10-138-145-301

Query Match          2.3%; Score 51.8; DB 6; Length 381;
Best Local Similarity 56.9%; Pred. No. 3.5;
Matches 95; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1051 ctgaagctttccaaacccaaaatgggaaagccgctctggagcgtgtaaacctcgacaca 1110
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 ctgattgtcccaagagactcaggaagatcctaataacacacctgttcgctcgatcata 263
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1111 gcaaggggtgcaactgtaagcgtcagcgtcgtgctgaggaactactgtgagtgatgagg 1170
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 agagaggatgtaattgcaagaagtcagctgtctcagaagaattgtcgaatgctatcagg 323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1171 ccaaaatcatgtgttcttcattgcataatgcattgttcgcaaaaac 1217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 gaggtgttggtggctcacaactgcagatgcgaaaacttgcaaaaac 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-10-027-632-221356
; Sequence 221356, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 221356
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-221356

Query Match          2.1%; Score 47.6; DB 6; Length 530;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 11:10:47 ; Search time 5416.12 Seconds
(without alignments)
8245.244 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aattcgggtcgaagcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

1	2127.6	99.7	2134	9	HSU86074	U86074 Homo sapien
2	915	42.9	157269	2	AP003096	AP003096 Homo sapi
3	672.2	31.5	2215	10	AB057423	AB057423 Mus muscu
4	662	31.0	1861	10	MMU77383	U77383 Mus musculu
5	647.8	30.4	2241	10	MMU67176	U67176 Mus musculu
6	208.6	9.8	157269	2	AP003096	AP003096 Homo sapi
7	152.8	7.2	188859	2	AP003732	AP003732 Homo sapi
8	139.6	6.5	22856	10	AB057422	AB057422 Mus muscu
9	122.4	5.7	188859	2	AP003732	AP003732 Homo sapi
10	122.2	5.7	179953	2	AC097959	AC097959 Rattus no
11	108.4	5.1	40893	3	CBRG45E19	AC084631 Caenorhab
12	102	4.8	39553	3	CEJCH	Z82274 Caenorhabdi
13	102	4.8	330724	2	CEY67H2	AL022475 Caenorhab
14	87.4	4.1	179953	2	AC097959	AC097959 Rattus no
15	83	3.9	214055	2	AC099385	AC099385 Rattus no
16	83	3.9	247253	2	AC103272	AC103272 Rattus no
17	83	3.9	297288	2	AC108533	AC108533 Rattus no
18	76.4	3.6	94362	8	AC005405	AC005405 BAC F18A1
19	76.4	3.6	170020	8	AP003256	AP003256 Oryza sat
20	76.4	3.6	180206	2	AP003274	AP003274 Oryza sat
21	74.6	3.5	187360	9	AC021105	AC021105 Homo sapi
22	73.6	3.4	97714	8	ATF25024	AL078469 Arabidops
23	73.6	3.4	101679	8	ATF19B15	AL078470 Arabidops
24	73.6	3.4	198429	8	ATCHRIV70	AL161574 Arabidops
25	71.2	3.3	100047	8	AC006081	AC006081 Arabidops
26	69.8	3.3	64971	2	AC107744	AC107744 Mus muscu
27	69.6	3.3	128463	2	AP003619	AP003619 Oryza sat
28	69.6	3.3	159947	2	AP003763	AP003763 Oryza sat
29	69.2	3.2	27555	2	AC012903	AC012903 Drosophil
30	69.2	3.2	146153	3	AC007452	AC007452 Drosophil
31	69.2	3.2	159065	3	AC091634	AC091634 Drosophil
32	69.2	3.2	264646	3	AE003818	AE003818 Drosophil
33	65	3.0	188844	2	AC099384	AC099384 Rattus no
34	57.4	2.7	157285	9	AC068483	AC068483 Homo sapi
35	56.8	2.7	1793	8	AY046019	AY046019 Arabidops
36	56.8	2.7	2298	8	AF204059	AF204059 Arabidops
37	56.8	2.7	2431	8	AF206324	AF206324 Arabidops
38	55.2	2.6	2146	8	AF205142	AF205142 Arabidops
39	55	2.6	188844	2	AC099384	AC099384 Rattus no
40	53.2	2.5	2153	9	AB049880	AB049880 Macaca fa
41	53.2	2.5	2296	9	HSM802712	AL359601 Homo sapi
42	53.2	2.5	125020	9	AF429315	AF429315 Homo sapi
43	52.8	2.5	13684	3	AE001403	AE001403 Plasmodi
44	50.8	2.4	253305	3	PFMAL3P7	AL034559 Plasmodi
45	50.4	2.4	368	6	AX284242	AX284242 Sequence

ALIGNMENTS

RESULT	1	HSU86074	2134 bp	mRNA	linear	PRI 13-APR-1999
LOCUS	HSU86074	Homo sapiens tesmin mRNA, complete cds.				
DEFINITION	U86074					
ACCESSION	U86074.1	GI:4581562				
VERSION						
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		1 (bases 1 to 2134)				
TITLE		Sugihara,T., Wadhwa,R., Kaul,S.C. and Mitsui,Y.				
JOURNAL		A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation				
MEDLINE		Genomics 57 (1), 130-136 (1999)				
REFERENCE		99208669				
AUTHORS		2 (bases 1 to 2134)				
TITLE		Sugihara,T.				
JOURNAL		Direct Submission				
TITLE		Submitted (20-JAN-1997) NIBH, AIST, Higashil-1, Tsukuba, Ibaraki				
FEATURES		305, Japan				
		Location/Qualifiers				

source 1..2134
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 CDS 407..1306
 /codon_start=1
 /product="tesmin"
 /protein_id="AAD24668.1"
 /db_xref="GI:4581563"

/translation="MVICQLKGGTOMLIDNSRTELKALHLVPOYQDNVNLQSDVP
 KPMFALRFELPASPKNLITQOLEGALPSVNGSAPSTLPGPPKITIAGYCDGF
 ASGDFCNMCNCCNLLHDIERPKAIKACLRNPEAFQPKIKGLGNVKKPOHNKG
 CNRRSGKLNKYCEYEAQIMCSSLCKICGNYEESPERKTLMSNPIMOTGLEGSS
 HYLPTKFSGLFRSHDRSPSCISWEVEATCACLLAQGEAEKEHCKSLAEOMIL
 BEFGRLSQILLHEFKSKLME"

BASE COUNT 586 a 490 c 528 g 530 t
 ORIGIN

Query Match 99.7%; Score 2127.6; DB 9; Length 2134;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2130; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	aattcggggtcaaggcgaagctgcgggggagacagcagcggggagctcctcgagg	60
DB	1	AATTCGGGCTCAAGGCGAAGCTCGCGGGGGCGACAGCGACGGGGAGCTCCTCGGGG	60
QY	61	agtaccccgagatccagagctcagcgctgaggaagctcgcgctcctgcagggccccc	120
DB	61	AGTACCCCGGATCCAGAGCTCAGCGCGCTGGAGAGCTCGCGCTCCTGCAGGCCCGC	120
QY	121	agccgcccgcctgcaacgtgcaactcctctcgtcgtgtacccgcgcacccgcagccgc	180
DB	121	AGCGCGCGCGCTCAAGGTGCACTCTCTGCTCCTGCTGTACCGCGCACCGCAGCGCGC	180
QY	181	gggtgtttgcccctggggcgctgggtcctggaaggagctccacccggcgctccgc	240
DB	181	GGGTGTGTTCGCGCTGGCGCGCTGGGTCTGCGAAGGAGCCCTCCACCGCGCGCTCCG	240
QY	241	atgctccagttgaatcgaagtaagcaggtggtactactacaagtaataatccggaaga	300
DB	241	ATGATCCCACTTGAATCAAGGTAGCAGGTGTACTACTACAAGTAATAATCCGGAAGA	300
QY	301	agcaactttgcagaatcttcttcaggaatacctgttgcaagtctccatggtccacgga	360
DB	301	AGCAACTTTGCAGAACTCTTCTCAGGAATCTCTTGAAGTTCCTCATGTCCTCCAGGA	360
QY	361	actagagatgcctcgtctgtcttctaagaagattccaccccaatggtgatgcc	420
DB	361	ACTAGAGGATGCCTCTCTGCTCTTAAAGAAAGATTCCACCCCAATGTGTATGCCA	420
QY	421	atgaaagggggacacaaatgctatgtagacaattcttagaacaagagaactaaagc	480
DB	421	ATTGAAAGGGGGACACAAATGCTATGTATAGACAATCTAGAACAGAGACAATAAAGC	480
QY	481	actccattggttctcagtcatacgaatacaataataatctacagtcagatgtccctaa	540
DB	481	ACTCAATTTGCTCAGTATCAAGATCAAAATAAATATATCTACAGTCAGATGTCCCTAA	540
QY	541	accaatgactgcttagtaggagattttgcccagatcaacaaattaaatctcattac	600
DB	541	ACCAATGACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAACAAATAAATCTCATTAC	600
QY	601	acaacaacttgagggagcctaccatcggtagtcacacggctgtcttcccccgggac	660
DB	601	ACACAACTTGAGGAGCCTTACCATCGGTAGTCAACGGGTCTGCTTTCCCGCTCGGATC	660
QY	661	aactcttcaggaccacccaaaaataacttgggtggtgactgtgactgttggcagtg	720
DB	661	AACCTTTCCAGGACCACAAAAATAAATTTGGCTGGGTACTGTGACTGCTTTGGCACTGG	720
QY	721	ggacttttgcaaacactgcaattgtaataattgttgcaaacacttgcattgatattga	780
DB	721	GGACTTTTGCAACAACTGCAATTGTAATAATTGTGCAACAACTGCAATCATGATATGA	780

QY	781	acggtttaaagccattaaggcatgtcttgtagaataccagaagctttccagccaaaaat	840
DB	781	ACGGTTTAAAGCCATTAAAGGCATGTCTTGGTAGAATCCAGAAAGCTTTCCAGCCAAAAAT	840
QY	841	tgggaaggggcgaatggcaatgtcaagcccccagcacacaagaaggtgcaactgcagag	900
DB	841	TGGGAAGGGCCCAATTTGGGCAATGTCAAGCCCCAGCACACAAGGGTGCACACTGCAGGAG	900
QY	901	gtcaggctgcctgaagaattactgcagtgctatgagggcccaaatatgtgtctctctat	960
DB	901	GTCAAGCTGCCTGAAGAATTACTGCGAGTGTCTATGAGGCCCAAAATATATGTCTTCTAT	960
QY	961	ttgcataatgcatttggttgcataaatttatgaagaagccagacaacaaacactaataag	1020
DB	961	TTGCAAAATGCATTGGTTGCAAAATTTATGAAGAAAGCCAGAACGAAACACTAATATAG	1020
QY	1021	catgcaaaactacatgcagactggaggtttggaaggcagccattcctcccccacacgaa	1080
DB	1021	CATGCCAAACTACATGCAGACTGGAGTTTGAAGCCAGCCATTACTGCCACCAACGAA	1080
QY	1081	atttcagagacttccaagattcagtcacgatagggcgccctcctcatgcatctcctggga	1140
DB	1081	ATTTTCAGGACTTCCAAGATTCAAGTTCAGTCACGATAGGCGGCTTCTCATGATCTCTCTG	1140
QY	1141	ggtggtggaggccacatgcgctgcctgtctcagggagagaagcccgagaaagaaaca	1200
DB	1141	GGTGTGGAGGCGACATGCGCCTGCTGCTCAGGAGAGAGAGCCCGAGAAAGAAACA	1200
QY	1201	ctgctccaaagtcctggcagagcagatgatcctggaggaatttgggaaggtgcttatcaca	1260
DB	1201	CTGCTCCAAAGTGCCTGGCAGAGCAGATGATCTGAGGAATTTGGGAAGGTGCTTATCACA	1260
QY	1261	gattctccacactgagtttaaatcttaaggattgaaattggagtagatataaagtgtga	1320
DB	1261	GATTCTCCACACTGAGTTTAAATCTAAGGATTTGAAATGGAGTAGATATAAAGTGTGA	1320
QY	1321	atgcatgttgatttcttagcttagaatactctagtttagaagaagatgttaggggaa	1380
DB	1321	ATGCATGTTGATTGCTTCTTAGTCTAGAAATCTCTAGTTTAGAAAGATGTTTAGGGGAA	1380
QY	1381	catgagctggctctgcagcaacacagcctccctgcatccctggcccgccagggagttt	1440
DB	1381	CATGAGCTGGCTCTGCAGCAACACAGCTCCCTCGCATCCCTGGCCCGCCAGGATTT	1440
QY	1441	actcagagctctctgaagatgtggcaacccatcccccctttctgaggaagtcagtgcc	1500
DB	1441	ACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCCCTTTTCTGAGGAGGTGCATGGCC	1500
QY	1501	tgagcattgttctgcccagagagagagcttggtttcccatagctcctgggagagtg	1560
DB	1501	TGAGCATTTGTTCTCTGGCCCCAGGAGAGAGCTTGGGTTCCCATAGTCTCTGGAGAGTG	1560
QY	1561	tctgcaggcgcgagggcgagagcagcctgcggagagctcactctgctgactcttc	1620
DB	1561	TCGTCAGGGCGCGGAGGCGAGAGAGCCCTCGGGAGAGCTCACCTCTGCTGACTCTTC	1620
QY	1621	ctctcagagaatgtctctgagggtctctgcatagaacccctaatggtttcttctgtt	1680
DB	1621	CTCTCAGAGAATGTGCTCTGGAGGCTGCTCTGCATGAAACCCCTAATGTGTTCTTCTTT	1680
QY	1681	gttttccaaattatttagaaaaaagttctccggatgggtgtgtgtatataccacttaaat	1740
DB	1681	GTTTTTCAAATTTATGAAAAAAGTTCTCCGGATGGGCTGTTGTGTATACCACTTAANAT	1740
QY	1741	ctctagaagaactactgaacacccctaaagatttctgtagcgtagatatatttccccagagaca	1800
DB	1741	CTCTAGAGAACTACTGAACACCTAAAGATTTCTGTAGCGTAGATATTTCCCCCAGAGACA	1800
QY	1801	cgcgaactgtcagctcttcttaaggcccccgagagacgcagggcaatggtggtcctgcagggc	1860
DB	1801	CGCGAACTGTCTTCTTAAGGCCCGCCCGGAGAGCGCAGGCAATGGGCTCTCGCAGGC	1860

```

QY 1861 caggcttcaccagcatgtcttgagtagagactaaataattaccagttctctctgtgt 1920
      |||
Db 1861 CAGGCTTCACCAGCAGTCTTTGAGTTAGAGGACTTAAATATATCCAGTTCTCTGTGT 1920
QY 1921 ttctactgaattgtggaagaactctattatccaattaaactctctccataattattgtgt 1980
      |||
Db 1921 TTCTACTTGAATGTGGAAGAAGCTTATTATCCAAATTAACCTTCTCCATAAATATTGTGT 1980
QY 1981 aatattattattgtttgaaacatggttcacataactagctgttggaacaccagcagga 2040
      |||
Db 1981 AATATATATTATTGTTGTTGTTAAACATGGTTCACATACTAGCTTGTGGAACACCAGCAGGA 2040
QY 2041 aaatgaattcttaagtgcgctttttgttctgttgaagcaaaagatgaataaaattt 2100
      |||
Db 2041 AAATGAATTTCTTAAAGTTGACGCTTTGTTGTTCTGTTGTAAGCAAGATCAATAAAAAATTT 2100
QY 2101 ccaatgtcgaataaaataaaataaaataaaataaaataaa 2134
      |||
Db 2101 CCAATGTGCAAAAAAATAAATAAATAAATAAATAAATAA 2134

RESULT 2
AP003096/c
LOCUS
DEFINITION
  AP003096 157269 bp DNA linear HTG 18-JAN-2001
  Homo sapiens chromosome 11 clone CTD-2007L18 map 11q, WORKING DRAFT
  SEQUENCE, 32 unordered pieces.
ACCESSION
  AP003096
VERSION
  HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
  Homo sapiens DNA, clone:CTD-2007L18.
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 157269)
AUTHORS
  Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
  Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  Homo sapiens 157,269 genomic DNA of 11q
  Published Only in Database (2001) In press
  2 (bases 1 to 157269)
AUTHORS
  Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
  Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  Direct Submission
  Submitted (16-JAN-2001) Masahira Hattori, The Institute of Physical
  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
  1-7-22 Suohiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
  (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
  Tel:81-45-503-9111, Fax:81-45-503-9170)
  ----- Genomic Center
  Center: RIKEN Genomic Sciences Center(GSC)
  Center code: RIKEN
  Web site: http://hgp.gsc.riken.go.jp/
  Contact: hattori@psc.riken.go.jp
  ----- Project Information
  Center project name: HumDrafitl
  Center clone name: CTD-2007L18
  ----- Summary Statistics
  Sequencing vector: PCR products; 100% of reads
  Chemistry: Dye-terminator ET-amersham; 100% of reads
  Assembly program: Phrap; version 0.990329
  Consensus quality: 147566 bases at least Q40
  Consensus quality: 151779 bases at least Q30
  Consensus quality: 153479 bases at least Q20
  Insert size: 154169; sum-of-contigs
  Quality coverage: 8.32x in Q20 bases; sum-of-contigs
  -----
  NOTE: This is a 'working draft' sequence. It currently consists of
  32 contigs. The true order of the pieces is not known and their
  order in this sequence record is arbitrary. Gaps between the
  contigs are represented as runs N, but the exact sizes of the gaps
  are unknown. This record will be updated with the finished sequence
  as soon as it is available and the accession number will be
  preserved
  1 11499 contig of 11499 bp in length

```

```

      11600 19042 contig of 7443 bp in length
      19143 26427 contig of 7285 bp in length
      26528 38740 contig of 12213 bp in length
      38841 55483 contig of 16643 bp in length
      67548 74728 contig of 11965 bp in length
      74829 79575 contig of 4747 bp in length
      79676 88957 contig of 9282 bp in length
      89058 93854 contig of 4797 bp in length
      93955 102673 contig of 3659 bp in length
      99015 105297 contig of 2524 bp in length
      102774 110398 contig of 4980 bp in length
      110478 117540 contig of 7063 bp in length
      117641 120176 contig of 2536 bp in length
      120277 125707 contig of 5431 bp in length
      125808 128247 contig of 2440 bp in length
      128348 131115 contig of 2768 bp in length
      131216 134071 contig of 2856 bp in length
      134172 136288 contig of 2117 bp in length
      136389 139198 contig of 2810 bp in length
      139299 141579 contig of 2281 bp in length
      141680 143825 contig of 2146 bp in length
      143926 146191 contig of 2266 bp in length
      146292 148382 contig of 2091 bp in length
      148483 149666 contig of 1184 bp in length
      149767 151747 contig of 1981 bp in length
      151848 153751 contig of 1904 bp in length
      153852 154877 contig of 1026 bp in length
      154978 156158 contig of 1181 bp in length
      156259 157269 contig of 1011 bp in length.
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 32 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will
  * be preserved.
  * 1 11499: contig of 11499 bp in length
  * 11500 11599: gap of 100 bp
  * 11600 19042: contig of 7443 bp in length
  * 19043 19142: gap of 100 bp
  * 19143 26427: contig of 7285 bp in length
  * 26428 26527: gap of 100 bp
  * 26528 38740: contig of 12213 bp in length
  * 38741 38840: gap of 100 bp
  * 38841 55483: contig of 16643 bp in length
  * 55484 55583: gap of 100 bp
  * 55584 67548: contig of 11965 bp in length
  * 67549 74728: contig of 7080 bp in length
  * 74729 74828: gap of 100 bp
  * 74829 79575: contig of 4747 bp in length
  * 79576 79675: gap of 100 bp
  * 79676 88957: contig of 9282 bp in length
  * 88958 89057: gap of 100 bp
  * 89058 93854: contig of 4797 bp in length
  * 93855 93954: gap of 100 bp
  * 93955 98914: contig of 4960 bp in length
  * 98915 99014: gap of 100 bp
  * 99015 102673: contig of 3659 bp in length
  * 102674 102773: gap of 100 bp
  * 102774 105297: contig of 2524 bp in length
  * 105298 105397: gap of 100 bp
  * 105398 110377: contig of 4980 bp in length
  * 110378 110477: gap of 100 bp
  * 110478 117540: contig of 7063 bp in length
  * 117541 117640: gap of 100 bp
  * 117641 120176: contig of 2536 bp in length
  * 120177 120276: gap of 100 bp
  * 120277 125707: contig of 5431 bp in length
  * 125708 125807: gap of 100 bp

```


RESULT	5
MMU67176	
LOCUS	
DEFINITION	2241 bp mRNA linear
ACCESSION	Mus musculus tesmin-1 mRNA, complete cds.
VERSION	U67176
	U67176.1 GI:4581558

Web site: <http://hap.gsc.riken.go.jp/>
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11
 Center clone name: CTD-2007L18
 ----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 147566 bases at least Q40
 Consensus quality: 151779 bases at least Q30
 Consensus quality: 153479 bases at least Q20
 Insert size: 154169; sum-of-contigs
 Quality coverage: 8.32x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 11499 contig of 11499 bp in length
11600 13042 contig of 7443 bp in length
19143 26427 contig of 7285 bp in length
26528 38740 contig of 12213 bp in length
38841 55483 contig of 16643 bp in length
55584 67548 contig of 11965 bp in length
67649 74728 contig of 7080 bp in length
74829 79575 contig of 4747 bp in length
79676 88957 contig of 9282 bp in length
89058 93854 contig of 4960 bp in length
8915 98914 contig of 3659 bp in length
99015 102673 contig of 2524 bp in length
102774 105297 contig of 9282 bp in length
105398 110377 contig of 4980 bp in length
110478 117540 contig of 7063 bp in length
117641 120176 contig of 2536 bp in length
120277 125707 contig of 5431 bp in length
125808 128247 contig of 2440 bp in length
128248 128347 contig of 100 bp in length
128348 131115 contig of 2768 bp in length
131116 131215 contig of 100 bp in length
131216 134071 contig of 2856 bp in length
134072 134171 contig of 100 bp in length
134172 136288 contig of 2117 bp in length
136289 136388 contig of 100 bp in length
136389 139198 contig of 2810 bp in length
139199 139298 contig of 100 bp in length
139299 141579 contig of 2281 bp in length
141580 141679 contig of 100 bp in length
141680 143825 contig of 2146 bp in length
143826 143925 contig of 100 bp in length
143926 146191 contig of 2266 bp in length
146192 146291 contig of 100 bp in length
146292 148382 contig of 2091 bp in length
148383 148482 contig of 100 bp in length
148483 149666 contig of 1184 bp in length
149667 149766 contig of 100 bp in length
149767 151747 contig of 1981 bp in length
151748 151847 contig of 100 bp in length
151848 153751 contig of 1904 bp in length
153752 153851 contig of 1026 bp in length
153852 154877 contig of 106 bp in length
154878 156158 contig of 1181 bp in length
156159 156258 contig of 100 bp in length
156259 157269 contig of 1011 bp in length.
    
```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 32 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 11499: contig of 11499 bp in length
11500 11599: gap of 100 bp
11600 19042: contig of 7443 bp in length
19043 19142: gap of 100 bp
19143 26427: contig of 7285 bp in length
26428 26527: gap of 100 bp
26528 38740: contig of 12213 bp in length
38741 38840: gap of 100 bp
38841 55483: contig of 16643 bp in length
55484 55583: gap of 100 bp
55584 67548: contig of 11965 bp in length
    
```

```

* 67549 67648: gap of 100 bp
* 67649 74728: contig of 7080 bp in length
* 74729 74828: gap of 100 bp
* 74829 79575: contig of 4747 bp in length
* 79576 79675: gap of 100 bp
* 79676 88957: contig of 9282 bp in length
* 88958 89057: gap of 100 bp
* 89058 93854: contig of 4979 bp in length
* 93855 93954: gap of 100 bp
* 93955 98914: contig of 4960 bp in length
* 98915 99014: gap of 100 bp
* 99015 102673: contig of 3659 bp in length
* 102674 102773: gap of 100 bp
* 102774 105297: contig of 2524 bp in length
* 105298 105397: gap of 100 bp
* 105398 110377: contig of 4980 bp in length
* 110378 110477: gap of 100 bp
* 110478 117540: contig of 7063 bp in length
* 117541 117640: gap of 100 bp
* 117641 120176: contig of 2536 bp in length
* 120177 120276: gap of 100 bp
* 120277 125707: contig of 5431 bp in length
* 125708 125807: gap of 100 bp
* 125808 128247: contig of 2440 bp in length
* 128248 128347: gap of 100 bp
* 128348 131115: contig of 2768 bp in length
* 131116 131215: gap of 100 bp
* 131216 134071: contig of 2856 bp in length
* 134072 134171: gap of 100 bp
* 134172 136288: contig of 2117 bp in length
* 136289 136388: gap of 100 bp
* 136389 139198: contig of 2810 bp in length
* 139199 139298: gap of 100 bp
* 139299 141579: contig of 2281 bp in length
* 141580 141679: gap of 100 bp
* 141680 143825: contig of 2146 bp in length
* 143826 143925: gap of 100 bp
* 143926 146191: contig of 2266 bp in length
* 146192 146291: gap of 100 bp
* 146292 148382: contig of 2091 bp in length
* 148383 148482: gap of 100 bp
* 148483 149666: contig of 1184 bp in length
* 149667 149766: gap of 100 bp
* 149767 151747: contig of 1981 bp in length
* 151748 151847: gap of 100 bp
* 151848 153751: contig of 1904 bp in length
* 153752 153851: gap of 100 bp
* 153852 154877: contig of 1026 bp in length
* 154878 156158: contig of 1181 bp in length
* 156159 156258: gap of 100 bp
* 156259 157269: contig of 1011 bp in length.
    
```

FEATURES

```

source
1..157269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CTD-2007L18"
/note="assembly_fragment"
11600..19042
/note="assembly_fragment"
19143..26427
/note="assembly_fragment"
26528..38740
/note="assembly_fragment"
38841..55483
/note="assembly_fragment"
55584..67548
/note="assembly_fragment"
67649..74728
/note="assembly_fragment"
    
```

```

misc_feature
1..11499
misc_feature
11600..19042
misc_feature
19143..26427
misc_feature
26528..38740
misc_feature
38841..55483
misc_feature
55584..67548
misc_feature
67649..74728
    
```

COMMENT

Tel:81-45-503-9111, Fax:81-45-503-9170)
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@hgp.gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11
 Center clone name: RP11-682D22
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 181083 bases at least Q40
 Consensus quality: 183471 bases at least Q30
 Consensus quality: 184447 bases at least Q20
 Insert size: 185059; sum-of-contigs
 Quality coverage: 7.99x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 39 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

19831 contig of 19831 bp in length
 19932 31497 contig of 11586 bp in length
 31598 44647 contig of 13050 bp in length
 44748 54198 contig of 9451 bp in length
 54299 63363 contig of 9065 bp in length
 63464 71678 contig of 8215 bp in length
 71779 79960 contig of 8182 bp in length
 80061 87061 contig of 7001 bp in length
 87162 93934 contig of 7673 bp in length
 94035 101099 contig of 7065 bp in length
 101200 108623 contig of 7424 bp in length
 108724 114505 contig of 5782 bp in length
 114606 119720 contig of 5115 bp in length
 119821 124688 contig of 4868 bp in length
 124789 130420 contig of 5632 bp in length
 130521 134923 contig of 4403 bp in length
 135024 138158 contig of 3135 bp in length
 138259 143086 contig of 4828 bp in length
 143187 146616 contig of 3430 bp in length
 146717 150135 contig of 3419 bp in length
 150236 153465 contig of 3230 bp in length
 153566 156263 contig of 2698 bp in length
 156364 158882 contig of 2519 bp in length
 158983 161109 contig of 2127 bp in length
 161210 164035 contig of 2826 bp in length
 164136 166612 contig of 2477 bp in length
 166713 168665 contig of 1953 bp in length
 168766 171614 contig of 2849 bp in length
 171715 173550 contig of 1836 bp in length
 173651 174290 contig of 640 bp in length
 174391 176225 contig of 1835 bp in length
 176326 178179 contig of 1834 bp in length
 178280 179640 contig of 1361 bp in length
 179741 181689 contig of 1949 bp in length
 181790 183060 contig of 1271 bp in length
 183161 184208 contig of 1048 bp in length
 184309 185664 contig of 1356 bp in length
 185765 187616 contig of 1852 bp in length
 187717 188859 contig of 1143 bp in length.

* NOTE: This is a 'working draft' sequence. It currently consists of 39 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

74829. .79575
 /note="assembly_fragment"
 79676. .88957
 /note="assembly_fragment"
 89058. .93854
 /note="assembly_fragment"
 93955. .98914
 /note="assembly_fragment"
 99015. .102673
 /note="assembly_fragment"
 102774. .105297
 /note="assembly_fragment"
 105398. .110377
 /note="assembly_fragment"
 110478. .117540
 /note="assembly_fragment"
 117641. .120176
 /note="assembly_fragment"
 120277. .125707
 /note="assembly_fragment"
 125808. .128247
 /note="assembly_fragment"
 128348. .131115
 /note="assembly_fragment"
 131216. .134071
 /note="assembly_fragment"
 134172. .136288

Query Match 9.88; Score 208.6; DB 2; Length 157269;
 Best Local Similarity 94.8%; Pred. No. 2.8e-34;
 Matches 238; Conservative 0; Mismatches 9; Indels 4; Gaps 2;
 Qy 8 ggtcaaggcagctcggggggcagcagcggggagctctctcggggagtaacc 67
 Db 120578 GGTCAAGCGAAGCTCGGGGGGGCGAGCGCGGGAGCTCTCGGGGAGTACCC 120637
 Qy 68 cgggacccagagctcagcgctggagagcgtcgctctcgcagcccgagccgcc 127
 Db 120638 CGGGATCCACAGCTACGCGCGTGGAGGAGCTCGCGCTCTCGAGGCCCGCAGCCGCC 120697
 Qy 128 cgcctgaacgtgaactctctctcgtctgctaccgcgcagcagccgcgggtgtt 187
 Db 120698 CGCCTGCAACGTGACTTCTCTGCTGCTACTACCGCGCAGCGCGCGGGGTGT 120757
 Qy 188 ttgccccctggcctggctcgcagagagcctccaccggcgctcgcgatgacc 247
 Db 120758 ---GCCCTGGCGCCTGGTGGTCTG- GAAGGAGCTCCACCGGGCGTCCGATGATCC 120813
 Qy 248 cagtgaaatc 258
 Db 120814 CAGTAAGAAC 120824

RESULT 7
 LOCUS AP003732 188859 bp DNA linear HTG 12-JUN-2001
 DEFINITION Homo sapiens chromosome 11 clone RP11-682D22 map 11q, WORKING DRAFT
 SEQUENCE, 39 unordered pieces.
 ACCESSION AP003732
 VERSION AP003732.1 GI:14349298
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens DNA, clone:RP11-682D22.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
 Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (11-JUN-2001) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@hgp.gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/)

FEATURES	Source	Location/Qualifiers
*	185665	185764: gap of 100 bp
*	185765	187616: contig of 1852 bp in length
*	187617	187716: gap of 100 bp
*	187717	188859: contig of 1143 bp in length.
		Location/Qualifiers
	1.	.188859
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="11"	
	/map="11q"	
	/clone="RP11-682D22"	
	/note="assembly_fragment"	
misc_feature	1932..31497	
misc_feature	/note="assembly_fragment"	
misc_feature	31598..44647	
misc_feature	/note="assembly_fragment"	
misc_feature	44748..54198	
misc_feature	/note="assembly_fragment"	
misc_feature	54299..63363	
misc_feature	/note="assembly_fragment"	
misc_feature	63464..71678	
misc_feature	/note="assembly_fragment"	
misc_feature	71779..79960	
misc_feature	/note="assembly_fragment"	
misc_feature	80061..87061	
misc_feature	/note="assembly_fragment"	
misc_feature	87162..93934	
misc_feature	/note="assembly_fragment"	
misc_feature	94035..101099	
misc_feature	/note="assembly_fragment"	
misc_feature	101200..108623	
misc_feature	/note="assembly_fragment"	
misc_feature	108724..114505	
misc_feature	/note="assembly_fragment"	
misc_feature	114606..119720	
Query Match		
Best Local Similarity 7.2%; Score 152.8; DB 2; Length 188859;		
Matches 166; Conservative 0; Mismatches 22; Indels 0; Gaps 0;		
QY	239	gcattgccccagttgaaatcaagtgtaagcaggtggtactactacaagtcaataatcccgaa 298
Db	141012	GCTTTTTC TAGTGGTAATCAAGACGAGCGTGGTACTACTACANGTAATAATCCGGAA 141071
QY	299	gaagcaactttgcagaatcttctgtcaggaatccctgttgcagattcccatggtccacg 358
Db	141072	GAACCACTTTG CAGAACTCTTTGCTCAGGAATCCTGTTCAAAGTTCCTCATCGTCCAG 141131
QY	359	gaactagagatgcctcctgctgtctcttctaagaagaatcccaacccaatggtgatatgc 418
Db	141132	GAACTAGAGATGCTCTGCTGTCTCTTTAAGAAAGATTCCAAACCAATGGTAATTATG 141191
QY	419	caattgaa 426
Db	141192	CTAAAGAA 141199
RESULT 8		
AB057422	AB057422	22856 bp DNA linear ROD 26-DEC-2001
LOCUS	Mus musculus	gene for tesmin, complete cds, strain:129.
DEFINITION	AB057422	
ACCESSION	AB057422	
VERSION	AB057422.1	GI:15617419
KEYWORDS		
SOURCE	Mus musculus (strain:129)	DNA.
ORGANISM	Mus musculus	
REFERENCE		
AUTHORS		
TITLE		

[illegible]

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 179953)

REFERENCE AUTHORS

1 (bases 1 to 17953)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Allbrooks, S.L., Anaratunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Blmage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowle, S., Brivia, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draeger, H., Dugan-Rocha, S., Durbun, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Gaspar, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsgree, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Nassey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, K., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission

JOURNAL
REFERENCE
2 (bases 1 to 179953)

AUTHORS
Worley, K. C.

TITLE
JOURNAL.

COMMENT

Submitted (23-Oct-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16901822.

Center: Baylor College of Medicine
Center code: BCM
Web Site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GFSK
----- Summary Statistics -----
Center clone name: CH230-114G10
Assembly program: phrap; version 0.990329First call to

Consensus quality: 163172 bases at least Q40
Consensus quality: 168807 bases at least Q30
Consensus quality: 174078 bases at least Q20
Estimated insert size: 160479; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

- * NOTE: Estimated insert size may differ from sequence length
- * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
- * NOTE: This is a 'working draft' sequence. It currently

* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	6639:	contig of 6639 bp in length
*		6739:	gap of unknown length
*	6640	12683:	contig of 5944 bp in length
*	6740	12684:	gap of unknown length
*	12684	16867:	contig of 4084 bp in length
*	12784	16867:	gap of unknown length
*	16868	22929:	contig of 5962 bp in length
*	16968	23029:	gap of unknown length
*	22930	30265:	contig of 7236 bp in length
*	23030	30365:	gap of unknown length
*	30266	30366:	contig of 6024 bp in length
*	30366	36489:	gap of unknown length
*	36390	42068:	contig of 5579 bp in length
*	36490	42168:	gap of unknown length
*	42069	46728:	contig of 4560 bp in length
*	42169	46828:	gap of unknown length
*	46729	52208:	contig of 5380 bp in length
*	46829	52308:	gap of unknown length
*	52209	56609:	contig of 4301 bp in length
*	52309	56709:	gap of unknown length
*	56610	59548:	contig of 2839 bp in length
*	56710	59648:	gap of unknown length
*	59549	62710:	contig of 3082 bp in length
*	59649	62811:	gap of unknown length
*	62711	66327:	contig of 3417 bp in length
*	62811	66328:	gap of unknown length
*	66328	69089:	contig of 2762 bp in length
*	66329	69189:	gap of unknown length
*	69090	72645:	contig of 3456 bp in length
*	69190	72745:	gap of unknown length
*	72846	76817:	contig of 4072 bp in length
*	72746	76917:	gap of unknown length
*	76818	80140:	contig of 3123 bp in length
*	76918	80140:	gap of unknown length
*	80041	83579:	contig of 3439 bp in length
*	80141	83679:	gap of unknown length
*	83580	84646:	contig of 2786 bp in length
*	83680	86466:	gap of unknown length
*	86466	90515:	contig of 3950 bp in length
*	86566	90615:	gap of unknown length
*	90516	94222:	contig of 3607 bp in length
*	90616	94322:	gap of unknown length
*	94223	97924:	contig of 3602 bp in length
*	94323	98024:	gap of unknown length
*	97925	98025:	contig of 1727 bp in length
*	98025	99851:	gap of unknown length
*	99851	102643:	contig of 2791 bp in length
*	102643	102743:	gap of unknown length
*	102743	106793:	contig of 4056 bp in length
*	106793	106898:	gap of unknown length
*	106899	110343:	contig of 3445 bp in length
*	110344	110443:	gap of unknown length
*	110444	112934:	contig of 2491 bp in length
*	112934	112935:	gap of unknown length
*	112935	116034:	contig of 3000 bp in length
*	116035	116134:	gap of unknown length
*	116135	119554:	contig of 3420 bp in length
*	119555	119654:	gap of unknown length
*	119655	120209:	contig of 2375 bp in length
*	120210	122129:	gap of unknown length
*	122130	125046:	contig of 2917 bp in length
*	122130	125047:	gap of unknown length
*	125047	127444:	contig of 2298 bp in length
*	127445	127544:	gap of unknown length
*	127545	129251:	contig of 1707 bp in length
*	129252	129351:	gap of unknown length
*	129352		

```

* 129352 131551: contig of 2200 bp in length
* 131551 131651: gap of unknown length
* 131651 131716: contig of 2065 bp in length
* 131716 133816: gap of unknown length
* 133816 136695: contig of 2879 bp in length
* 136695 136795: gap of unknown length
* 136795 138444: contig of 1649 bp in length
* 138444 138544: gap of unknown length
* 138544 140364: contig of 1820 bp in length
* 140364 140464: gap of unknown length
* 140464 142056: contig of 1592 bp in length
* 142056 142156: gap of unknown length
* 142156 144300: contig of 2144 bp in length
* 144300 144400: gap of unknown length
* 144400 145630: contig of 1230 bp in length
* 145630 145730: gap of unknown length
* 145730 147365: contig of 1835 bp in length
* 147365 147665: gap of unknown length
* 147665 149525: contig of 1860 bp in length
* 149525 149625: gap of unknown length
* 149625 151295: contig of 1670 bp in length
* 151295 151395: gap of unknown length
* 151395 152702: contig of 1307 bp in length
* 152702 152802: gap of unknown length
* 152802 154292: contig of 1490 bp in length
* 154292 154392: gap of unknown length
* 154392 155990: contig of 1598 bp in length
* 155990 156090: gap of unknown length
* 156090 157685: contig of 1595 bp in length
* 157685 157785: gap of unknown length
* 157785 159784: contig of 1999 bp in length
* 159784 159884: gap of unknown length
* 159884 161347: contig of 1463 bp in length
* 161347 161447: gap of unknown length
* 161447 162678: contig of 1231 bp in length
* 162678 162778: gap of unknown length
* 162778 163996: contig of 1218 bp in length
* 163996 164096: gap of unknown length
* 164096 166978: contig of 2882 bp in length
* 166978 167078: gap of unknown length
* 167078 168859: contig of 1781 bp in length
* 168859 168959: gap of unknown length
* 168959 170817: contig of 1858 bp in length
* 170817 170917: gap of unknown length
* 170917 172118: contig of 1101 bp in length
* 172118 172019: gap of unknown length

Query Match 5.7%; Score 122.2; DB 2; Length 179953;
Best Local Similarity 71.7%; Pred. No. 1.6e-15;
Matches 160; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 892 ctgcaggagctcagctgctgaagattactcgcgagtgctatgagcccaaatatgtg 951
Db 58954 CTGTGGGGGAAGCCTGATTTTCATCAAGCATTTTATCATCAGCGCCAAATACAGTG 59013

QY 952 ttctctatttgcacattgctgtgcaaaattatgaagaagcccaagaagac 1011
Db 59014 TTCTTCCATTGCAATGCTTTCGCAAAATCTATGAAGAAGCCAGAGCGAAAAAT 59073

QY 1012 actaatgagcatgccaaactacatcagactgaggttttgaagcagcattacatgcc 1071
Db 59074 GCTGATGAGCAGCCCGCTTACATGAGCGCTTGGGAGCTTTGAGAGCAGCCACCATCTGTC 59133

QY 1072 accaagcaattttcaggacttccaagattcagtcagatag 1114
Db 59134 CCCAGCAAAATTCAGGAGCTCCAGACTGAGAAAAATAGG 59176

RESULT 11
CBRG45E19 40893 bp DNA linear INV 04-NOV-2000
LOCUS Caenorhabditis briggsae cosmid G45E19, complete sequence.
DEFINITION
ACCESSION AC084631

```

```

VERSION AC084631.1 GI:11095080
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 40893)
AUTHORS Washington University Genome Sequencing Center.
TITLE The C. briggsae Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40893)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplie@watson.wustl.edu

FEATURES
source
1..40893
/organism="Caenorhabditis briggsae"
/strain="GujArat G16"
/db_xref="taxon:6238"
/clone="G45E19"

BASE COUNT 13086 a 7791 c 7927 g 12089 t
ORIGIN

Query Match 5.1%; Score 108.4; DB 3; Length 40893;
Best Local Similarity 62.7%; Pred. No. 1.1e-12;
Matches 188; Conservative 0; Mismatches 106; Indels 6; Gaps 1;

QY 698 tactgtgactgttccagtgaggactttgcaacaactgcaattgtaattgtgc 757
Db 38000 TACTGTGACTGCTTCCGGAACGGAGAGAGTTTTCGCGTACTGCAACTGCAGGACTGCCAC 38059

QY 758 acaactgcatcatgatattgaacgggtttaagccattaaagcattgtcttggtagaaat 817
Db 38060 AACATATAGATTATGACAGCCCAACGATCAAAAGCCATTTCGTGAGTCTTTGGAGCGAAAT 38119

QY 818 ccagaagctttccagcccaaaattgggaaggcccaattgggca-----atgtcaagccc 871
Db 38120 CCAAAATGCTTTCAAGCCGAAATTCGGAATAGCTCGTGGTGAACCCGCGGATATTGAACGC 38179

QY 872 cagcaacaagaagggtgcaactgcaggaggtcaggtcgtcgtgaagaattactcagatgc 931
Db 38180 TTGCATCAAAAGGGATGCCACTGCACAGAAAAGCGGTTCCTGAAAACATATTTCGAGTGC 38239

QY 932 tatgagggcccaattatgtgttcttctatttgcacaaatgcattggttgcacaaatgatga 991
Db 38240 TAGGAAGCAAAAGTGCCTGTACTGATCGATGCAAGTGAAGAGCTTCAGAAATACGGAA 38299

RESULT 12
CEJCB8
LOCUS Caenorhabditis elegans cosmid Jc8, complete sequence.
DEFINITION
ACCESSION Z82274
VERSION Z82274.1 GI:2814248
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (sites)

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

AUTHORS	note.	
	Genome sequence of the nematode <i>C. elegans</i> : a platform for investigating biology. The <i>C. elegans</i> Sequencing Consortium Science 282 (5396), 2012-2018 (1998)	
JOURNAL	99069613	
MEDLINE		
REFERENCE	2 (bases 1 to 39553)	
AUTHORS	Lightning, J.	
	Direct Submission	
JOURNAL	Submitted (11-NOV-1996) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jesus@sanger.ac.uk or rwenematode.wustl.edu	
	On Jan 28, 1998 this sequence version replaced gi:2558532.	
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.	
	Current sequence finishing criteria for the <i>C. elegans</i> genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.	
	This sequence is the entire insert of clone JC8.	
	The true right end of clone C48D1 is at 6889 in this sequence. The start of this sequence (1..6889) overlaps with the end of sequence Z81049.	
FEATURES	The end of this sequence (39450..39553) overlaps with the start of sequence AL32951.	
	For a graphical representation of this sequence and its analysis see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=JC8	
source	IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.	
	Location/Qualifiers	
gene	1..39553	
	/organism="Caenorhabditis elegans"	
CDS	/db_xref="taxon:6239"	
	/chromosome="IV"	
	/clone="JC8"	
	join(4308..4352,4737..4953,5892..5902)	
	/gene="JC8.11"	
	join(4308..4352,4737..4953,5892..5902)	
	/gene="JC8.11"	
	/note="cDNA EST yk245a7.3 comes from this gene"	
	cDNA EST yk245a7.5 comes from this gene"	
	/codon_start=1	
	/protein_id="CAB05233.2"	
	/db_xref="GI:14530485"	
	/db_xref="SPTREMBL:O62296"	
	/translation="MGSWNYSIGTWFGEVDENLQKHEDSLRVETASMDQLQKIRDDSYKRAMGEIDPIELPFRDRDSSISIPACVINTVESQTPIRHLINAD"	
gene	complement(join(7174..7377,8392..9150,10178..10645))	
	/gene="JC8.2"	
CDS	complement(join(7174..7377,8392..9150,10178..10645))	
	/gene="JC8.2"	
	/note="Contains similarity to Pfam domain: PF00400 (WD domain, G-beta repeat), Score=44.5, E-value=7.6e-10, N=2"	
	cDNA EST yk329e7.3 comes from this gene	
	cDNA EST yk404a8.3 comes from this gene	
	cDNA EST yk329e7.5 comes from this gene	
	cDNA EST yk404a8.5 comes from this gene"	
	/codon_start=1	
	/protein_id="CAB05230.2"	
	/db_xref="GI:6425181"	
	/db_xref="SPTREMBL:Q9NAQ2"	
	/translation="MKTPKPELMDGLVMSVDSLSIQRGVAKEIPDKIKLDEKLKELLEGAPLSTONETELDDDESDNEKLGRLRGDVEKMDSEFEKKYMKGYNEIKDEPQCQMGKAMTSTKDDPYVTEQVDSDEEEKDEIMVKDDNNMVAVAKIDKGDFTLECVYNEADSWYCHHDIYILDAPPLCIEPQHPDNGEETGKGNLLAVGTMNSIHIWDLQIMTATFTFLTGKKERYKGAARKRONSAQCHTDVAISLAWNRITTHVLASGGA DKTIVLWDLDEAKPAQIIPQGGEGIQTMKWHNPNESTFLLLTGTMKGQVNVVDCRESSGN	

ASAAKFDGQIEKVIWHPNFATFACSSDDGRLRLHLDMRKPGCELGWEGVAHDGPIGL
 TISALTKGLVTVGEDMMVWKVVEDTNGGIEKVHSEKLTIGELHCAQFNPDAVLS
 VGTAAADLIRIIDLTKYEPVVKAFSE"
 complement(join(11334..11829,12629..12735,12794..12924,
 13189..13450))
 /gene="JC8.13"
 complement(join(11334..11829,12629..12735,12794..12924,
 13189..13450))
 /gene="JC8.13"
 /codon_start=1
 /protein_id="CAD21646.1"
 /db_xref="GI:18369715"
 /translation="MASCWCPVAEINFRSRNHLIDQLEATSTSSGPRSKYVADPT
 MKVEYSRSAAADTHKYKPELLRPFPVLOHTIDYLLDYSPEKDRRAITSKOPASIF
 SVFSDRLRSRQDMIMQNLRGDFTVILLIEKMLPYIETDCACKMAKQCSVDPLKHDQ
 LECFGKWFEEVNSNDIIPNLISACFFRQLHRKSSILHDLFLRRKKLPKPSFLNLI
 QSISSFSYNNRRFFDFOOLOPLKHSISDCVSLIRKQMSNISVAFPTPVARLPS
 QLLSDWLGFPANLEFFDFLRYSVIDEQCNLILSALRPVVEISYQDFSSROYE"
 complement(join(13717..13814,13864..13971,14182..14436,
 14562..14598))
 /gene="JC8.3a"
 complement(join(13717..13814,13864..13971,14182..14436,
 14562..14598))
 /gene="JC8.3a"
 /note="predicted using GeneFinder
 contains similarity to Pfam domain: PF00298 (Ribosomal
 protein L11), Score=176.7, E-value=1.2e-49, N=1
 cDNA EST EMBL:R00775 comes from this gene; cDNA EST
 yk90d2.5 comes from this gene; cDNA EST yk139d10.3 comes
 from this gene; cDNA EST yk139d10.5 comes from this gene
 cDNA EST yk122f7.5 comes from this gene; cDNA EST
 yk149h7.5 comes from this gene; cDNA EST yk151h12.5 comes
 from this gene; cDNA EST yk165e9.5 comes from this gene
 cDNA EST yk149h7.3 comes from this gene; cDNA EST
 yk213e10.3 comes from this gene; cDNA EST yk236f7.3 comes
 from this gene; cDNA EST yk241b1.3 comes from this gene
 cDNA EST yk254h9.3 comes from this gene; cDNA EST
 yk268g12.3 comes from this gene; cDNA EST yk269g3.3 comes
 from this gene; cDNA EST yk273e7.3 comes from this gene
 cDNA EST yk278d11.3 comes from this gene; cDNA EST
 yk284f2.3 comes from this gene; cDNA EST yk335c10.3 comes
 from this gene; cDNA EST yk335g12.3 comes from this gene
 cDNA EST yk299a2.3 comes from this gene; cDNA EST
 yk372f10.3 comes from this gene; cDNA EST yk391h1.3 comes
 from this gene; cDNA EST yk374g7.3 comes from this gene
 cDNA EST yk376a3.3 comes from this gene; cDNA EST
 yk402b6.3 comes from this gene; cDNA EST yk417g2.3 comes
 from this gene; cDNA EST yk425f6.3 comes from this gene
 cDNA EST yk448e7.3 comes from this gene; cDNA EST
 yk452b3.3 comes from this gene; cDNA EST yk452d7.3 comes
 from this gene; cDNA EST yk474a2.3 comes from this gene
 cDNA EST yk464d9.3 comes from this gene; cDNA EST
 yk409h4.3 comes from this gene; cDNA EST yk485e8.3 comes
 from this gene; cDNA EST yk485e8.3 comes from this gene
 cDNA EST yk446g3.3 comes from this gene; cDNA EST
 yk475e10.3 comes from this gene; cDNA EST yk483g3.3 comes
 from this gene; cDNA EST yk501g7.3 comes from this gene
 cDNA EST yk505g10.3 comes from this gene; cDNA EST
 yk501c3.3 comes from this gene; cDNA EST yk505d6.3 comes
 from this gene; cDNA EST yk213e10.5 comes from this gene
 cDNA EST yk236f7.5 comes from this gene; cDNA EST
 yk241b1.5 comes from this gene; cDNA EST yk254h9.5 comes
 from this gene; cDNA EST yk268g12.5 comes from this gene
 cDNA EST yk278d11.5 comes from this gene; cDNA EST
 yk269g3.5 comes from this gene; cDNA EST yk273e7.5 comes
 from this gene; cDNA EST yk284f2.5 comes from this gene
 cDNA EST yk372a2.5 comes from this gene; cDNA EST
 yk335g12.5 comes from this gene; cDNA EST yk299a2.5 comes
 from this gene; cDNA EST yk372f10.5 comes from this gene
 cDNA EST yk391h1.5 comes from this gene; cDNA EST
 yk374g7.5 comes from this gene; cDNA EST yk376a3.5 comes
 from this gene; cDNA EST yk402b6.5 comes from this gene
 cDNA EST yk417g2.5 comes from this gene; cDNA EST

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 179953)
 Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Blmaga,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle-McD., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlson,E., Kelly,S., Khan,U., King,L., Korwah,J., Kowar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sison,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczyk,R., Woodden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 179953)
 Worley,K.C.

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:16901822.

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GFSK
 Center clone name: CH230-114G10
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to findPhrapList
 Consensus quality: 163172 bases at least Q40
 Consensus quality: 168807 bases at least Q30
 Consensus quality: 174078 bases at least Q20
 Estimated insert size: 160479; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 62 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 6639: contig of 6639 bp in length
 6739: gap of unknown length
 12683: contig of 5944 bp in length
 12684: gap of unknown length
 12784: contig of 4084 bp in length
 16868: gap of unknown length
 16867: contig of 5962 bp in length
 22929: contig of 5962 bp in length
 22930: gap of unknown length
 30265: contig of 7236 bp in length
 30266: gap of unknown length
 30365: contig of 6024 bp in length
 30366: gap of unknown length
 36390: contig of 5579 bp in length
 36490: gap of unknown length
 42069: contig of 4560 bp in length
 42169: gap of unknown length
 46729: contig of 5380 bp in length
 46829: gap of unknown length
 52209: contig of 4301 bp in length
 52309: gap of unknown length
 56610: contig of 2839 bp in length
 56710: gap of unknown length
 59549: contig of 3062 bp in length
 59649: gap of unknown length
 62711: contig of 3417 bp in length
 62810: gap of unknown length
 66227: contig of 2762 bp in length
 66328: gap of unknown length
 69090: contig of 3456 bp in length
 69190: gap of unknown length
 72646: contig of 4072 bp in length
 72745: gap of unknown length
 76817: contig of 3123 bp in length
 76918: gap of unknown length
 80041: contig of 3439 bp in length
 80141: gap of unknown length
 83579: contig of 2786 bp in length
 83680: gap of unknown length
 86465: contig of 3950 bp in length
 86565: gap of unknown length
 90515: contig of 3607 bp in length
 90615: gap of unknown length
 94222: contig of 3602 bp in length
 94223: gap of unknown length
 94323: contig of 1727 bp in length
 97925: gap of unknown length
 98025: contig of 2791 bp in length
 98252: gap of unknown length
 102642: contig of 4056 bp in length
 102743: gap of unknown length
 106798: contig of 3445 bp in length
 106799: gap of unknown length
 106899: contig of 2491 bp in length
 110344: gap of unknown length
 110344: contig of 3000 bp in length
 112935: gap of unknown length
 116035: contig of 3420 bp in length
 116135: gap of unknown length
 119555: contig of 2375 bp in length
 119555: gap of unknown length
 122029: contig of 2917 bp in length
 122129: gap of unknown length
 125046: contig of 2298 bp in length
 125146: gap of unknown length
 125147: contig of 1707 bp in length
 127445: gap of unknown length
 127545: contig of 1707 bp in length
 129251: gap of unknown length
 129252: gap of unknown length

GenCore version 4.5
Copyright (c) 1993 - 2000. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 10:54:07 ; Search time 479.19 Seconds
(without alignments)
7646.018 Million cell updates/sec.

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aattcggggtcaggcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2134	100.0	2134	21	AA288156 Human testis speci
2	662	31.0	1861	21	AA288157 Mouse testis speci
3	647.8	30.4	2241	21	AA288155 Mouse testis speci
4	476.8	22.3	3495	22	AAK94268 Human full-length
5	429	20.1	445	22	AA101300 Human reproductive
6	245	11.5	250	22	AAH35563 Human colon cancer
7	203.8	9.6	3533	20	AAV72867 Human lin-54 homol
8	199.8	9.4	588	22	AAK91874 Human cDNA 5'-end
9	199.8	9.4	588	22	AAK93281 Human cDNA clone r

10	147.4	6.9	1093	23	AA82145	DNA encoding novel
11	113	5.3	3164	23	ABL10437	Drosophila melanog
12	102	4.8	1503	20	AAV72865	Caenorhabditis ele
13	69.2	3.2	8372	23	ABL10436	Drosophila melanog
14	52	2.4	986	20	AAK51735	DNA encoding a hum
15	52	2.4	1772	21	AAF21809	Human breast and o
16	50.4	2.4	368	22	AAAG6046	Human cancer agent
17	50.4	2.4	3413	21	AAA47459	Human TANGO 239 co
18	50.4	2.4	3413	21	AAA47483	Human TANGO 239 co
19	50	2.3	3234	13	AAQ30999	Notch clone hN3k f
20	48.6	2.3	2028	20	AAK81861	CDNA encoding a mu
21	48.6	2.3	2028	21	AAAO8010	Mouse polynucleoti
22	48.2	2.3	12405	22	AAAS5330	Chemically pretrea
23	48.2	2.3	12405	24	AAAS61143	Human gene regulat
24	48	2.2	207	21	AAAC98751	Human colon cancer
25	48	2.2	4100	20	AAZ32019	Human METH1 relate
26	48	2.2	4100	22	AAAC90076	LI3855 CDNA clone.
27	47.6	2.2	2607	22	AAH33503	Human colon cancer
28	47.4	2.2	1909	20	AAV84356	Nuclear matrix-ass
29	47.4	2.2	1912	14	AAQ43443	AML 1 gene. Homo
30	47.4	2.2	4193	23	AA572413	DNA encoding novel
31	47.4	2.2	4287	19	AAV20475	Human AML1/MTG8 on
32	47.4	2.2	5828	20	AAH84592	AML1-MTG16 fusion
33	47.4	2.2	6056	20	AAH84593	AML1-MTG16 fusion
34	47	2.2	3035	24	ABK09774	Human ovarian tumo
35	46.8	2.2	1034	21	AAZ52527	Human secreted pro
36	46.8	2.2	1181	19	AAV59803	Human secreted pro
37	46.8	2.2	1212	19	AAV59686	Human secreted pro
38	46.6	2.2	10640	22	AAO37729	P. faiciparum telo
39	46.4	2.2	378	22	AAI90863	Human polynucleoti
40	46.4	2.2	6695	22	AAK70340	Human immune/haema
41	46.4	2.2	6695	22	AAK82282	Human immune/haema
42	46.2	2.2	1705	20	AAV08856	Gene encoding huma
43	46.2	2.2	1705	20	AAV84616	Human secreted pro
44	46.2	2.2	1705	22	ABA83399	Human secreted pro
45	46	2.2	768	22	AAI96382	Human neuroblastom

ALIGNMENTS

RESULT 1

AA288156
ID AA288156 standard; cDNA; 2134 BP.

XX AC AA288156;

XX DT 25-APR-2000 (first entry)

XX DE Human testis specific factor tesmin encoding cDNA SEQ ID NO:3.

XX KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
XX KW differentiation regulatory factor; male germ cell regulatory actor;
XX KW germ cell differentiation; sterility; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 407..1306

XX FT FT /*tag= a

XX FT FT /product= "tesmin"

XX PN WO200004147-A1.

XX PD 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-JP03859.

XX PR 17-JUL-1998; 98JP-0219856.

XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

Qy 1861 caggcttgaccagcagctgctgagtagagcacttaaaattaccagttcttctgtgt 1920
 Db 1861 caggcttgaccagcagctgctgagtagagcacttaaaattaccagttcttctgtgt 1920
 Qy 1921 ttctactgaatttgaaagcctctatttcccaataaactctccataattattgtgt 1980
 Db 1921 ttctactgaatttgaaagcctctatttcccaataaactctccataattattgtgt 1980
 Qy 1981 aatattatttcttgtaaaacatggttcacataactagcttctggaaccagcaggtta 2040
 Db 1981 aatattatttcttgtaaaacatggttcacataactagcttctggaaccagcaggtta 2040
 Qy 2041 aaatgaattcttaagtgcagcttctgttctgttgtaaaagcaaatgaataaaattt 2100
 Db 2041 aaatgaattcttaagtgcagcttctgttctgttgtaaaagcaaatgaataaaattt 2100
 Qy 2101 ccaatgtcgaataaaataaaataaaataaaataaaataaaataaaataaaataaa 2134
 Db 2101 ccaatgtcgaataaaataaaataaaataaaataaaataaaataaaataaaataaa 2134

RESULT 2

AAZ88157 standard; cDNA; 1861 BP.

AAZ88157;

25-APR-2000 (first entry)

Mouse testis specific factor tesmin encoding cDNA SEQ ID NO:2.

Testis specific factor; tesmin; cell death; regulation; spermatocyte; differentiation regulatory factor; male germ cell regulatory actor; germ cell differentiation; sterility; ss.

Mus musculus.

Key Location/Qualifiers
 CDS 271..1158
 FT /*tag= a
 FT /product= "tesmin"

W0200004147-A1.

27-JAN-2000.

16-JUL-1999; 99WO-JP03859.

17-JUL-1998; 98JP-0219856.

(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

WPI; 2000-147785/13.
 P-PSDB; AAY68463.

New male germ cell regulatory factor tesmin expressed in spermatocytes useful for investigation of germ cell differentiation and sterility -

Example 2; Page 38-42; 63pp; Japanese.

The present sequence encodes a male germ cell regulatory factor expressed specifically in spermatocytes, designated tesmin. Tesmin can be used in the investigation of the mechanisms of germ cell differentiation and sterility.

Sequence 1861 BP; 456 A; 481 C; 498 G; 426 T; 0 other;

Query Match 31.0%; Score 662; DB 21; Length 1861;
 Best Local Similarity 74.7%; Pred. No. 1.6e-147;

Matches 863; Conservative 0; Mismatches 280; Indels 13; Gaps 2;
 Qy 183 gtgttttgcctcctggcgccgtggtcctggaaggagcctccaccggcgctccgcat 242
 Db 48 gtcttacttcccggtgctgcggccgcgtgtgtccgtaagcgagagtcgcgt 107
 Qy 243 gatcccgatgaatacaaggttaagcaggtggtactactacacagtaataatccggaagaag 302
 Db 108 gatcccgatgaatacaag-aagcaggtggtgtagtggtccaggcgccgctgaagacg 166
 Qy 303 caatttgcagaaattcttctcaggaatcctgttgcgaatcccatcctcagggagac 362
 Db 167 caggttccaggccctcctggtcaggaatcctgttgcgaatcccatcctcagggag 226
 Qy 363 tagaggatgctcctgctgttcttcttaagaagaattcccaaccatggtgtagatgcgaat 422
 Db 227 cagaggagccctcagctgcctcgggaagaagactccagcccatggtgtagtgcgc 286
 Qy 423 tgaaggggggcacacaaatgctatgtatagacaaatcttagacacagagaacataaaagcac 482
 Db 287 tgaaggaggcgccagatgctgtcatagacaaactgtggtgcggaggagctcaaaagcgc 346
 Qy 483 tccatttggctcctcagatcaagatcaaaataattatctacagtcagatgcctcaaac 542
 Db 347 tccatttggctcctcagatcaagatcaaaataattatctacagtcagatgcctcaaac 406
 Qy 543 caatgactgctttagtaggagatttttgcagcatcaacaaataaataaattcattacac 602
 Db 407 caatgacaaactttagtaggagatttttgcagcatcaacaaataaataaattcattacac 466
 Qy 603 aacaaactttagtaggagatttttgcagcatcaacaaataaataaattcattacac 662
 Db 467 aggtgataatgagctcctccatcagctgtcaaatgggtcgtcttccctcctggaactg 526
 Qy 663 ctctccaggaccacccaaataaacttgggtggtgactgtgactgcttgcagtggtggg 722
 Db 527 ctctgcaaggcccccacaaataaacttgggtggtgactgtgactgcttccagcgggg 586
 Qy 723 acttttgcacaaactgcaattgtaataatttgcacaaacttgcacatcatatattgaac 782
 Db 587 acttctcaacagctgca-----gctgcacaaactgcgcctgagctcgagc 634
 Qy 783 ggtttaaagccattaaaggcagctgtcttgtagaataatccagagcttccagccaaaaattg 842
 Db 635 gctcaaaagccataaaaggcgtgtcttgtagaataatcctgagcttccacacaaaaatgg 694
 Qy 843 ggaaggcccaattgggcaattgtcaagccccagcaacaaagggtgcgaactgcaggaggt 902
 Db 695 ggaaggccgtctggagctgctaaacttcgacacagcaagggtgcgaactgtaagcgt 754
 Qy 903 caggctgcctgaagaattactgcagtgctgtagggcccaaatattgttcttctattt 962
 Db 755 caggctgcctgaagaattactgcagtgctgtagggcccaaatattgttcttctattt 814
 Qy 963 gcaaatgcatgtgtgcaaaaattatgaagaagcccaagacgaagaacacacataagca 1022
 Db 815 gcaaatgcatgtgtgcaaaaattatgaagaagcccaagacgaagaacacacataagca 874
 Qy 1023 tgcacaaactacatgcagactggaggttgggaaggcagcattacatgcacacacgaat 1082
 Db 875 caccctacatgcagactggaggttgggaaggcagcattacatgcacacacgaat 934
 Qy 1083 tttcaggacttccagattcagtcagtaggcgccttccctcctcctcctcctggagg 1142
 Db 935 tctcaggacctcccaaaactgagaaaaataggcaggcccttctcctcctcctcctggagg 994
 Qy 1143 tgggtggaggccacatgcgcctgcctgtcctcagggagagagggccgagaaagacact 1202
 Db 995 tagtggaggccacatgtgcctgcctgtgcggccagggtgaggaagcagagcagact 1054
 Qy 1203 gctccaaagtgcctggcagagcagatgatcctcggagggaatttggagggtgttattacaga 1262
 Db 1055 gtcccccaagcttggtgagcagatgatcctcggaggaggttggagggtgctctgcgcaga 1114

QY 1263 ttctccacactgagtttaaaatcgaaggattgaaatggagtagagtagatataaaagtgaat 1322
 Db 1115 ttctccacatcgagttcaagtcacaaagggtgaaattgagtagcgtgcaagctgtaaa 1174
 QY 1323 gcatgttgattgtc 1338
 Db 1175 ggggaaatgcctgtgc 1190

RESULT 3
 AA288155
 ID AA288155 standard; cDNA; 2241 BP.
 XX
 AC AA288155;
 XX
 XX 25-APR-2000 (first entry)
 DT
 XX Mouse testis specific factor tesmin encoding cDNA SEQ ID NO:1.
 DE
 XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility; ss.
 XX
 OS Mus musculus.
 XX
 XX Key Location/Qualifiers
 FH 651..1538
 FT CDS /tag= a
 FT /product= "tesmin"
 FT
 XX W0200004147-A1.
 XX
 XX 27-JAN-2000.
 XX
 PF 16-JUL-1999; 99WO-JP03859.
 XX
 PR 17-JUL-1998; 98JP-0219856.
 XX
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX
 XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX WPI; 2000-147785/13.
 DR P-PSDB; AAV68463.
 XX
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 XX
 PS Claim 3; Page 33-37; 63pp; Japanese.
 XX
 XX The present sequence encodes a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 2241 BP; 558 A; 554 C; 583 G; 546 T; 0 other;

Query Match 30.4%; Score 647.8; DB 21; Length 2241;
 Best Local Similarity 76.1%; Pred. NO. 4.1e-144;
 Matches 829; Conservative 0; Mismatches 247; Indels 13; Gaps 2;

QY 250 gttgaaatcaagtaagcaggtgtgactactacaagtaataatccggaagaagcaacttt 309
 Db 495 gttgaaatcaag-aagcaggtgtgagtgccaggcggcgcctggaagcagcagcttt 553
 QY 310 gcagaatctctgtcaggaatcctgttgcaagttcccatggtcccgaggaactagagga 369
 Db 554 ccaggccctctggtcaggaatcctgttgcaagttcccatcctccagagcagagga 613
 QY 370 tgctctctgctgtctctttaagaagaattcccaacccaatggtgatatgccaattgaaag 429

Db 614 ggctccagctgcctcgcgaagaagaactccagcccatggtgattgtcagctgaagg 673
 QY 430 gggcacacaaatgctatgtatagacaattctagaacaagagaactaaagcactccattt 489
 Db 674 aggcgccagatgctctgcatagacaactgtgagcgagagctcaagagcgtccatct 733
 QY 490 ggttctcagtatcaagatcaaaataatctacagtcagatgtccctaaacaaatgac 549
 Db 734 gcttctcagtagcagatgaccagagcagtttccctcagtcagagctccctaagcaatgac 793
 QY 550 tgcttagtaggagatatttggccagcatcaacaataataatctcattacacacaact 609
 Db 794 aacttttagggaagacttctgcagtagcagcgaagttaaatctcatcacaggtga 853
 QY 610 tgaggagccttaccatcgtagtaaacagggtctgttctccctcgggcatcaactctcc 669
 Db 854 taatgagctctccatcagctgcataggggctgtcttccctcgggactctctgca 913
 QY 670 aggacacacaaataaacttggctggtgactgactgcttggcagtgaggactttg 729
 Db 914 agggccacacaaataaactctgctggtgactgactgcttccagcgggactctg 973
 QY 730 caacaactgcaattgttaataattgttgcacaactgcatcatgatattgaacggttaa 789
 Db 974 caacagctgca-----gctgcacaacactgagcctgagcagcgtcaa 1021
 QY 790 agcattaaaggcatgtcttgtagaataccagaagcttccagcccaaaattgggaagg 849
 Db 1022 agccataaagggtgcttgatagaatacctgagcttcccaacccaataagggaagg 1081
 QY 850 ccaattgggcaatgtcaagcccgacacaaagggtgcaactgcagaggtcaggtg 909
 Db 1082 cgtctggagctgctaaacttcgacacagaaagggtgcaactgtaagcgtcaggctg 1141
 QY 910 cctgaagaattactgcagtgctatgagcccaaatatgttcttctatttgcgaattg 969
 Db 1142 cctgaagaactactgtgagtgctatgagcccaaatcatgttcttccatttgcaattg 1201
 QY 970 cattgttgcacaaatattgaagaagcccgacacgaagagacataatgagcatgcaaa 1029
 Db 1202 cattgttgcacaaactatgaagaaggtccagacgaataatgctgatgagcacacca 1261
 QY 1030 ctacatgcagactggaggttggaaaggcagccattaccctgccaccacgaattttcag 1089
 Db 1262 ctacatggagcctgggactttgagagcagccattattgtcccccagcaagtctcag 1321
 QY 1090 acttccaaattcagtcacagtagcggccttctcctcatgcatctcctggaggtggtga 1149
 Db 1322 acctcaaaactgagaaaaataggcagccttctctctgtatctctctgggaagttag 1381
 QY 1150 ggcacatgcgctcgtctgctcaggaagaaggcgcgaggaagaaacactgctccaa 1209
 Db 1382 ggcacatgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1441
 QY 1210 gtgcctggcagagcagatgactcctgaggaatttggaaaggtgcttcatcacagattcca 1269
 Db 1442 aagcttgctgagcagatgactcctgaggaatttggaaaggtgctcgcagattctcca 1501
 QY 1270 cactgatttaataatcaagggtggaattggaatgagtagagtagataaaagtgtgaatgatt 1329
 Db 1502 catcgagttcaagttcaagggtggaattggaatgagtagagtagataaaagtgtgaatgatt 1561
 QY 1330 gatttggc 1338
 Db 1562 gctgtggc 1570

RESULT 4
 AA288155
 ID AA288155 standard; cDNA; 2241 BP.
 XX
 AC AA288155;

PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	01-DEC-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	06-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition -

Claim 1: SEQ ID NO 1301; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a coding sequence of the invention.

Sequence 445 BP: 136 A: 105 C: 90 G: 111 T: 3 other;

Qy	228	ccggggcggtccgcatgatccaggttgaaatcaaggtaagcaggtggtactactacaagta	287
Db	2	ccggggcggtccgcatgatccaggttgaaatcaagg-aagcaggtggtactactacaagta	60
Qy	288	ataatccgggaagcaactttgcagaaattctttgctcaggaatcctgttgcgaattcc	347
Db	61	ataatccgggaagcaactttgcagaaattctttgctcaggaatcctgttgcgaattcc	120

Qy	348	catcgtcccaaggaactagaggatgcctcctgctgcttctcttaagaagattcccaaccaa	407
Db	121	catcgtcccaaggaactagaggatgcctcctgctgcttctcttaagaagattcccaaccaa	180
Qy	408	tgctgatgcacaattgaagggggcacacaatactatgtatagacaattctagaacaa	467
Db	181	tggcgatagcccaattgaaagggggcacacaatactatgtatagacaattctagaacaa	240
Qy	468	gagaactaaaagcactccatttgggttcctcagtcataagatcacaataaattatctacgt	527
Db	241	gagaactaaaagcactccatttgggttcctcagtcataagatcacaataaattatctacgt	300
Qy	528	cagatgtccctaaaccaatgactcttttagtagaggagatttttccagagcatcaacaaaat	587
Db	301	cagatgtccctaaaccaatgactcttttagtagaggagatttttccagagcatcaacaaaat	360
Qy	588	taaatctattacacacaacttgaggagccttaccatcgtagtcgaacgggtctgctt	647
Db	361	taaatctattacacacaacttgaggagccttaccatcgtagtcgaacgggtctgctt	420
Qy	648	tcctcctgggataactcttccagg	672
Db	421	tcctcctgggataactcttccagg	445
RESULT 6			
AAH35563			
ID	AAH35563 standard; cDNA; 250 BP.		
XX			
AC	AAH35563;		
XX			
DT	03-SEP-2001 (first entry)		
XX			
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:2645.		
XX			
KW	Human; colon cancer; colon cancer antigen; diagnosis: detection;		
KW	colorectal carcinoma; chromosome 11; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200122920-A2.		
XX			
PD	05-APR-2001.		
XX			
PF	28-SEP-2000; 2000WO-US26524.		
XX			
PR	29-SEP-1999; 99US-0157137.		
PR	03-NOV-1999; 99US-0163280.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX			
DR	WPI; 2001-235357/24.		
XX			
DR	P-PSDB; AAG76158.		
XX			
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,		
XX	useful for preventing, diagnosing and/or treating colorectal cancers -		
PS	Claim 1; Page 4257-4258; 9803pp; English.		
XX			
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon		
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where		
CC	the proteins are collectively known as colon cancer antigens. The colon		
CC	cancer antigens have cytostatic activity and can be used in gene		
CC	therapy and vaccine production. N and P may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate P		
CC	expression. For example, N and P may be used to treat disorders		
CC	associated with decreased expression by rectifying mutations or deletions		
CC	in a patient's genome that affect the activity of P by expressing		
CC	inactive proteins or to supplement the patients own production of P.		
CC	Additionally, N may be used to produce the colon cancer-associated Ps,		
CC	by inserting the nucleic acids into a host cell and culturing the cell		

Qy	675	cacacaaaataaactttgctgggtactgtgactgctgttgcagcagggtggaacttttgcacaa	734
Db	1908	caaaatcaactgtgttgaatatgtatttgaattgttttgcgaattggtgaaatttttgcacaa	1967
Qy	735	actgcaattgtaataattttgcaacaaacttgcatactgatgatgaacggttttaaagcca	794
Db	1968	actgcaattgtactaatgttacacaaatttgaacatacgaaaaagaaagcgaacaaacaa	2027
Qy	795	ttaaggcattgtcttggtagaataatcagaagctttccagcccaaaaatttgggaaggcccaat	854
Db	2028	taaaggcatgctgtacagaaatccagaaagcctttaagcctaagatagggaaggaagga	2087
Qy	855	tgggcaatgtcaagcccgacacacaaaagggtcgaactgcagaggtcagcgtgcctga	914
Db	2088	agggagaactctgacgtctcatagcaaaagggtgaaattgcaacgatacaggatgtctta	2147
Qy	915	agaattactcgagtgctatgaggcccaaatatgtgtcttctatttgcgaaatgcattg	974
Db	2148	aaaactactgtaatgctatgaggcaaaaataatgtgttcctcaatacgcaaatgtattg	2207
Qy	975	gtgcacaaaattatgaagaaagccacagaaagacacactaatatgacatgcga-----aa	1029
Db	2208	gctgaagaatttgaagaaagccggaaaggaagacattgatgtatttggcagatgcag	2267
Qy	1030	ctacatgcagactggaggtttggaaggcagccttaccttgcaccaaagcaattttcagg	1089
Db	2268	ctgaagttaagggttacagcaacaaacagcagcaagcaagaagtattcctctcaaatctcag	2327
Qy	1090	acttcc-----aaagattcagtaacagataggcggccttctctact	1127
Db	2328	actgttactaggccaacaccagcttttaaatagtggaaggcgaaataattgcatttcat	2387
Qy	1128	gcattctctgggaggtgtgtgaggccaatgcgctgtcctgtctcaggagagaagagg	1187
Db	2388	tgttaactaagggaantacttgaagccacnctgaattgcnctctgccagcagcagcagcag	2447
Qy	1188	ccgagaagaacactgctccaaatgctgtggcagacagatgatctctggaggaatttggaa	1247
Db	2448	cagacaagaaggaaaaatcaaaagcagcagcgaacgagatgatactctgaggaatncggac	2507
Qy	1248	gggtcttatcacagattctccacactgagtttaataactaa	1287
Db	2508	natcttgcattgattgtcatcaactctgcaggaagagcga	2547

```

RESULT      8
AAK91874
ID AAK91874 standard; cDNA; 588 BP.
XX
AC AAK91874;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human cDNA 5'-end sequence, SEQ ID NO: 334.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
XX Homo sapiens.
XX
XX EP1130094-A2.
XX
XX 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-0114089.
XX
XX 08-JUL-1999; 99JP-0194486.
XX
XX 11-JAN-2000; 2000JP-0118774.
XX
XX 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
PI
XX WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
XX
XX Claim 2; SEQ ID NO 334; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesising the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is the nucleotide
XX sequence of the 5'-end of a cDNA provided in the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
XX SQ Sequence 588 BP; 92 A; 224 C; 190 G; 78 T; 4 other;

Query Match          9.4%; Score 199.8; DB 22; Length 588;
Best Local Similarity 89.8%; Pred. No. 1e-37;
Matches 247; Conservative 0; Mismatches 21; Indels 7; Gaps 3;

Qy 8 ggtcaagcgcgaagctcgcgggggcagcagcgacgcggggagctctctcggggagtaacc 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 ggtcaagcgaanctcgcgggggcagcagcgacgcggggagctctcctcggggagtaacc 380

Qy 68 cgggatccagagctcaacgcgcgttgagagagtcgcgtctctgcagcccgagcccgagccgcc 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 cgggatccagagctcagcgcgttgagagagatctcgtctctgcagggcccgagcccgagccgcc 440

Qy 128 cgcctgcacgctgcacttcctgtctcgtctcgtctgcacccgcagcccgagcccggtgtt 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 cgcctgcaacgtgcacttcctgtctcgtctcgtctgcacccgcagcccgagcccggtgtt 500

Qy 188 ttgcccctggggcgccctgggtctctcgtcgaagagagctccaccacccgggctccgcagatccc 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 ---gccctctggcgcctgggtccttgaaa---agacctccacccgggctccgcagatccc 554

```


XX
SQ Sequence 3164 BP; 880 A; 865 C; 810 G; 609 T; 0 other;

Query Match 5.3%; Score 113; DB 23; Length 3164;
Best Local Similarity 63.0%; Pred. No. 9.3e-17;
Matches 192; Conservative 0; Mismatches 110; Indels 3; Gaps 1;

Qy 698 tactgtgaactgctttggccagtgaggacttttgcacaacactgcaaatgttaataattgttc 757
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2565 tactgcgattgctttgcaaacgcgagtttgtcaggactcacctgcgaagtgcctt 2624
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 758 acaacttgcacatgatattgaacggtttaaaagccattaaaggcatgtctttagtagaaat 817
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2625 aacatatggactacgaagtggagcgagcgtgtattccagctgctcgatcgtaac 2684
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 818 ccagaagctttcccagccaataattgggaagggccaattgggccaatgtcgaagccccagcac 877
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2685 cccagcgctttaaacaccaaaatttacggcaccaccaatttcagg---tgatatgcgtctgc 2741
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 878 acaaaagggtgcacctgcagaggtgcaggctgcctgcgaagaattactgcgagtgctatgag 937
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2742 acaaaagggtgcacctgcgaagaatcgggctgcctcaagaactattgtgagtgctatgag 2801
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 938 gcccaaatattgttcttctatttgcgaatgcattggttcgaaaaattatgaagaaagc 997
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2802 gcaaagatccctgctccagcatatgtaaatgctgtgggtgctgaaacatggaagaccgt 2861
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 998 ccaga 1002
|| ||
Db 2862 ccgga 2866

RESULT 12
AAV72865
ID AAV72865 standard; cdna; 1503 BP.
AC AAV72865;
DT 29-MAR-1999 (first entry)
DE Caenorhabditis elegans synMuv gene lin-54 cdna.
KW LIN-54; synthetic multivulvar; SynMuv; signal transduction;
KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
KW cell proliferation; gene therapy; ss.
OS Caenorhabditis elegans.
FH Key Location/Qualifiers
FT CDS 17..1333
FT /*tag= a
PN WO9854299-A1.
XX
PD 03-DEC-1998.
XX
PF 28-MAY-1998; 98WO-US11043.
XX
PR 28-MAY-1997; 97US-0047996.
XX
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Ceol C, Horvitz HR, Lu X;
XX
DR WPI: 1999-045362/04.
DR P-PDB; AA083392.
XX
Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
PT useful for treating diseases associated with altered levels of cell
PT proliferation, e.g. carcinomas
XX
PS Claim 10; Fig 14; 70pp; English.

XX	{(PEKE) PE CORP NY.
XX	PA
XX	Venter JC, Adams M, Li PWD, Myers EW;
XX	PI
XX	WPI; 2001-656860/75.
XX	P-PSDB; ABB66333.
XX	XX
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
PT	PT
XX	PS
XX	Claim 1; SEQ ID NO 25790; 2lpp + Sequence Listing; English. .
XX	XX
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC	sequences (ABU01840-ABU16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	XX
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences .
XX	XX
XX	Sequence 8372 BP; 2351 A; 1921 C;1986 G; 2114 T; 0 other;

Qy	698	tactgtgaactgcttttcagatcgagggaacttttgcacaacaactgcaattgttaataattgtgtgc	757
Db	587	tactgtgatgttttcgcaaatggaggttctgtcgtgactgcaattgcgaagatgtgtcac	646
Qy	758	aacaacttcgatcatgatattgaacggtttaaagccattaaaggcatgctcttgttgaagaat	817
Db	647	ataatatagaatcacgacagtcagcgctcaaaagccatccgtcagtcactgtgagcgaaat	706
Qy	818	ccagaagctttccaggccaaaattgggaaggcccaattggg-----caattgcgaagccc	871
Db	707	ccgaagcctttcaagccaaaattggtattgtctgtggaggtattaccgacatcgaacgt	766
Qy	872	cagcacacaacaagggtgcgaactgcaggaggtcagggtgcctgaagaatctactgcaggtgc	931
Db	767	cttcacagaaggaagtgcactgtaaaaagagtggtgtgtctgaaaaaacattgtgaggtg	826
Qy	932	tatgagggcccaaatattatgtgtttcttctatttgcaaatgcattggttgcaaaaattatgaa	991
Db	827	tatgaagcaaaaggttccgtgtaccgcatcatcgaatgtcgaagagtgatcagaatactactaa	886

RESULT	13	
ABL10436		
ID	ABL10436	standard; cdna; 8372 bp.
XX	XX	
AC	ABL10436;	
XX	XX	
DT	26-MAR-2002	(first entry)
XX	XX	
DE	Drosophila	melanogaster expressed polynucleotide SEQ ID NO 25790.
XX	XX	
KW	Drosophila;	developmental biology; cell signalling; insecticide;
KW	pharmaceutical;	gene; ss.
XX	XX	
OS	Drosophila	melanogaster.
XX	XX	
PN	WO200171042-A2.	
XX	XX	
PD	27-SEP-2001.	
XX	XX	
PF	23-MAR-2001;	2001WO-US09231.
XX	XX	
PR	23-MAR-2000;	2000US-191637P.
PR	11-JUL-2000;	2000US-0614150.
PR	XX	

[illegible]

Db 1653 catggcaaaaaaaaaaaaaaaaaaaaaa 1684

Search completed: July 11, 2002, 10:54:47
Job time: 9604 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 09:12:55 ; Search time 3455.78 Seconds
(without alignments)
8334.588 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aattcgggtcaggcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	655.4	30.7	1730	11 AK015724	Mus muscu
2	545.4	25.6	842	10 B1829802	B1829802 603079876
3	513.4	24.1	518	9 A1874249	A1874249 tz63d10.x
4	500.6	23.5	507	9 AW293246	AW293246 UI-H-BI2-
5	498.6	23.4	505	10 B1711510	B1711510 id97g07.x
6	480.4	22.5	487	9 AW236823	AW236823 xm49e06.x
7	448.4	21.0	453	9 A1652317	A1652317 wb60a08.x
8	443.6	20.8	473	9 AW102615	AW102615 xd68h02.x
9	428.6	20.1	435	10 B1792942	B1792942 ie49e01.x
10	417.4	19.6	426	10 B1711797	B1711797 id97g07.y
11	413.8	19.4	429	9 AW629183	AW629183 h152b01.y
12	413.2	19.4	429	9 AA846474	AA846474 aj56d11.s
13	412	19.3	474	9 A1624835	A1624835 ts71g00.x
14	412	19.3	509	9 A1968123	A1968123 wu3d06.x
15	412	19.3	642	9 A1968415	A1968415 wu02h05.x
16	388.8	18.2	414	9 AA994165	AA994165 ot61a06.s
17	382.2	17.9	398	9 AA758741	AA758741 ah80b05.s

C 18	371.4	17.4	772	10 B1520646	B1520646 603071608
19	363.2	17.0	888	10 BG773009	BG773009 602721268
20	348.6	16.3	355	9 A1796344	A1796344 w19n08.x
21	342.2	16.0	363	9 A1538405	A1538405 tp53g06.x
22	341.4	16.0	346	10 BE501038	BE501038 7a36a03.x
23	332	15.6	333	9 AL040101	AL040101 DKF2p434C
24	310.4	14.5	442	9 A1307797	A1307797 tb28c05.x
25	291.8	13.7	505	10 BG088866	BG088866 H3158G12-
26	291.2	13.6	336	10 B1793224	B1793224 ie49e01.y
27	281.6	13.2	2010	11 AK015732	AK015732 Mus muscu
28	275.2	12.9	280	9 A1377830	A1377830 te58d11.x
29	273.6	12.8	280	9 AW511334	AW511334 hd45e04.x
30	267.8	12.5	420	10 BG272409	BG272409 nah30a05
31	267.2	12.5	280	9 A1827726	A1827726 wf11d07.x
32	259.2	12.1	264	9 AA903758	AA903758 ok66a02.s
33	254.8	11.9	258	9 AA970002	AA970002 op61d08.s
34	251.8	11.8	266	9 AF012383	AF012383 AF012383
35	221.4	10.4	223	10 T19339	T19339 cl2012t Tes
36	214	10.0	697	9 BB611478	BB611478 BB611478
37	208	9.7	312	9 AW270088	AW270088 xv37g12.x
38	190.8	8.9	574	9 A1981460	A1981460 pat.pk005
39	189.4	8.9	475	10 B1706999	B1706999 fq13f03.y
40	185.2	8.7	627	10 BG772890	BG772890 602721090
41	170.2	8.0	397	10 BF937310	BF937310 fm63c07.y
42	161.6	7.6	638	10 BF426275	BF426275 df69g02.y
43	159.6	7.5	768	10 BG977029	BG977029 602845588
44	157.6	7.4	171	9 AL040102	AL040102 DKF2p434C
45	150.2	7.0	704	9 AV715308	AV715308 AV715308

ALIGNMENTS

RESULT 1

AK015724	AK015724	1730 bp	mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930507A18.metallothionein-like 5, testis-specific (tesmin), full insert sequence.	linear	HTC 19-JAN-2002
LOCUS	AK015724				
DEFINITION	AK015724				
ACCESSION	AK015724				
VERSION	AK015724.1	GI:12854168			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus				
REFERENCE	1 (sites)				
AUTHORS	Carninci,P. and Hayashizaki,Y.				
JOURNAL	High-efficiency full-length cDNA cloning				
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)				
PUBMED	99279253				
REFERENCE	2 (sites)				
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipillar sequencer				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				


```

/db_xref="taxon:9606"
/clone="IMAGE:2293267"
/clone_lib="NCI_CGAP_Ov35"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; This library represents the normalized
version of NCI_CGAP_Ov23. Cloned unidirectionally.
Primer: Oligo dT. Average insert size 0.86 Kb. Tumor
types include: mixed Mullerian tumor, papillary serous,
clear cell, spindle cell. All are primary tumors,
metastasis positive. Constructed by Life Technologies."
BASE COUNT 169 a 106 c 103 g 140 t
ORIGIN

Query Match 24.1%; Score 513.4; DB 9; Length 518;
Best Local Similarity 99.8%; Pred. No. 1.7e-67;
Matches 514; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1594 cggagagctcaactgtgtgactcttctctcagagaatgttctctggaggtctgtctg 1653
|||||
Db 518 CGGAGAGCTCACTCTGGTCAGCTCTCCCTCAGAGAAATGTTCTCTGGAGGCTGCTCTG 459

QY 1654 catgaaccctaatggtttctgttggtttttccaaattatttagaataaagttctccgg 1713
|||||
Db 458 CATGAAACCCCTAATGTTCTCTGTTGTTGTTTTCAAATATTAGAAATAAGTTCTCCGG 399

QY 1714 atgggtgtgtataccacttaaaatctctagagaactactgaacacctaaagatttc 1773
|||||
Db 398 ATGGGCTGTTGTATACCACTTAAATCTCTAGAGAACTACTGAACACCTTAAAGATTTC 339

QY 1774 ttagcgtatattttccagagacacgcgaactgtcagcttcttcttaagggccccggg 1833
|||||
Db 338 TGTAGCTAGATATTTCCCGAGGACGCGCACTGCTCACTCTTCTTAAGGCCCGCGG 279

QY 1834 agacgagcgaatggggcctcgcagccagcgttgcaccagcatgtctttagttagtagga 1893
|||||
Db 278 AGACGAGGCAATGGGGCTTCGAGGCGCAGGCTTGACACGATGCTTGTAGTTAGAGGA 219

QY 1894 cttaaaattaccagttcttctgttcttctactgaattgtgaaagctctattatcc 1953
|||||
Db 218 CTTAAATATATCCAGTTCTCTGTTGTTTCTACTTGAATTTGTGAAAGCTCTATTATCC 159

QY 1954 aattaacttccataattattgttgaataattattattgtttgttaaaacatggttcaaca 2013
|||||
Db 158 AATTAATCTCCATATATTGTTGTAATATTATTATTCTTTGTAAACATGTTTCAACA 99

QY 2014 taactagctgtggaaccgcggtaaaatgaattttagttgacgcttttggtctg 2073
|||||
Db 98 TAAGTAGCTGTGTGAACACGAGGTAAATGAATTTCTTAAGTTGAGGCTTTTGGTTCTG 39

QY 2074 ttgtaagcaagatgaataaaatttccaatctc 2108
|||||
Db 38 TTGTAAAGCAAGATGAATAAAATTTCCAATGTC 4

RESULT 4
AW293246/c 507 bp mRNA linear EST 16-JAN-2000
LOCUS
DEFINITION
UI-H-B12-ahl-a-10-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
IMAGE:2727067 3', mRNA sequence.
ACCESSION
AW293246
VERSION
AW293246.1 GI:6699968
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 507)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

JOURNAL COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

Location/Qualifiers
1..507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2727067"
/clone_lib="NCI_CGAP_Sub4"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI_CGAP_Sub4 library is a subtracted library derived from
the NCI_CGAP_Sub2 library which is a subtracted library
derived from the NCI_CGAP_Sub1 library, which is a
subtracted library derived from BI. BI constitutes a
mixture of 21 normalized or subtracted NCI_CGAP
libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28,
NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5,
NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11,
NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_GLL1,
NCI_CGAP_Lu24, NCI_CGAP_Brn23, NCI_CGAP_Lu5
NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 1322376-1323911,
1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 :
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clones
1323912-1325831, 1471368-1472903, 1492104-1493255)
NCI_CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE
Clones 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clones 1257096-1258631, 1469064-1470983, 1475592-1476743
) NCI_CGAP_Pr22 pool 1 : LLAM 2457-2459, 2758-2759,
3062-3068 (IMAGE Clones 985608-986759, 1101192-1101959,
1217928-1220615) NCI_CGAP_Co10 pool 1 : LLAM 2644-2653,
2871-2872 (IMAGE Clones 1057416-1061255, 1144584-1145351
) Subtraction was performed as previously described
[Nonaldo, Lennon & Soares (1996): Normalization and
Subtraction: Two Approaches To Facilitate Gene Discovery.
Genome Research 6, 791-806.]
TAG_LIB=NCI_CGAP_Lu5
TAG_TISSUE=Lung
TAG_SEQ=CAAC

BASE COUNT 162 a 99 c 93 g 153 t
ORIGIN

Query Match 23.5%; Score 500.6; DB 9; Length 507;
Best Local Similarity 99.2%; Pred. No. 1.4e-65;
Matches 503; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1622 tctcagagaatgttgcctctgaggtgctctgcataaaacctaatggtttctgttg 1681
|||||
Db 507 TCTCAGAGAATGTTGCTCTCGAGGCTGCTCTGCATGAACCCCTAATGTTCTGTTG 448
QY 1682 ttttcaaatatttagaataaagttctccggtggtgtgtgtgtatcaccactaaatc 1741
|||||
Db 447 TTTTCAAAATTTAGAAATAAGTTCTCCGGATGGGCTTGTGTATACCACTTAAATC 388
QY 1742 tctagagaactactgaacacctaaagattttctgtagctagatatttccccagagacac 1801

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bernaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 457.

FEATURES

source
1. .487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2687554"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 1475992-1476743). Subtraction by Bento Soares and M. Fatima Bernaldo. "
BASE COUNT 160 a 99 c 92 g 136 t
ORIGIN

Query Match 22.5%; Score 480.4; DB 9; Length 487;
Best Local Similarity 99.8%; Pred. No. 1.5e-62;
Matches 481; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1627 gagaatgttctctgaggtgctctgcagtgaaacccctaatggtttctgtttttt 1686
Db 486 GAGAATGTTCTCTGGAGGCTGCTGTCATGAAACCTAATGTTCTGTTGTTT 427
QY 1687 caaattattgaaataagttcccgatggctgtgtgtagaccactaaatctctag 1746
Db 426 CAAATATTATGAAATPAGTTCCTGGATGGCTGTGTGATACCTTAAATCTCTAG 367
QY 1747 agaactactgaacacccataagattttctgtagcgttagatatattccccagacacgcgaa 1806
Db 366 AGAACTACTGAACACCTAAAGATTTCGTAGCGTAGATATTTCCCCAGAGCGACGAA 307
QY 1807 ctgtcagttcttcttaagcccccgggagacgcagcgaatggggcctcgagccaggct 1866
Db 306 CTGTCAGTCTTCTTAAGGCCCGGGAGACCGAGGCAATGGGGCCCTCGCAGGCCAGGCT 247
QY 1867 tgaccagcatgtcttgagtagagacttaaaattatccagttcttctgtttctac 1926
Db 246 TGACACGATGCTTGAGTTAGAGACTTAAATATATCCAGTTCTCTGTGTTCTAC 187
QY 1927 ttgaattgtgaaagctctattccaatttaactcttctccatattattgttgtaatt 1986
Db 186 TTGAATTGTGAAAGCTCTATTATCCAAATTAATCTCTCCATAATATTGTTGTAATATT 127
QY 1987 attatgtttgtaaaacataggttcacataactagctgttggaacacagcaggttaaatga 2046
Db 126 ATTATTGTTTGTAAACATGGTTTACATACTAGCTTGTGGAACCCAGCGTAAATGA 67
QY 2047 attcttaagtgcgcttttggctctgttgtaagcaaaagatgaataaatttccaatg 2106
Db 66 ATTCTTAAGTTGACGCTTTTGGTCTGTTGTAAGCAAGATGAATATAAATTTCCATG 7
QY 2107 tc 2108
Db 6 TC 5
RESULT 7
AI652317/c

LOCUS AI652317 453 bp mRNA linear EST 17-DEC-1999
DEFINITION WB60a08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2310038 3', mRNA sequence.
ACCESSION AI652317
VERSION AI652317.1 GI:4736296
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 453)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bernaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 422 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source
1. .453
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2310038"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 1475992-1476743). Subtraction by Bento Soares and M. Fatima Bernaldo. "
BASE COUNT 151 a 88 c 85 g 129 t
ORIGIN

Query Match 21.0%; Score 448.4; DB 9; Length 453;
Best Local Similarity 99.8%; Pred. No. 9.2e-58;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1659 aaaccctaagtgttctgttttttcaattattagaataagttctcgagatggg 1718
Db 453 AAAACCCCTAATGTTCTGTTGTTTTCATATTTAGAAAATAGTTCTCCGGATGGG 394
QY 1719 ctgttgatcacactaaatctctagagactactgaacacactaaagattttctgtag 1778
Db 393 CTGTTGTGATACCACCTTAAATCTCTAGAGAACTACTGAAACCTTAAAGATTTCTGTAG 334
QY 1779 cgtgatattttccccagagacgcgaactgtcagttcttcttaagcccccgagacg 1838
Db 333 CGTAGATATTTCCCGAGGACGCGAAGCTCAGTCTTCTTAAAGCCCGCCGGAGACG 274
QY 1839 caggcaatggggcctcgagccaggctgtcacacagcatgtcttgagttagagacttaa 1898
Db 273 CAGGCAATGGGGCTCTCGACGCCAGGCTTCACACAGCATGCTGTGAGTTAGAGACTTAA 214
QY 1899 aattatccagtttctctgttcttacttgatgtgggaaaagctctattatccaatta 1958
Db 213 AATTATCCAGTTTCTGTTGTTTCTACTTGAATTGTGGAAAGAGCTCTATTATCCAATTA 154

```

Qy 1959 acttctccataattattgttgtaattattattgtttgttaaaacatggttcacataact 2018
|||||
Db 153 ACTTCTCCATAAATTATGTTGTAATATTATTGTTTGTAAACATGGTTCACATAACT 94
|||||
Qy 2019 agcttgtaggaacacagcagggttaaatgattcttaagttagcgcttttggttctgttcta 2078
|||||
Db 93 AGCTTGTGGAAACCAAGCAGGTAAGTAAATCTTAAAGTTGACGCTTTGGTTCCTGTTGTA 34
|||||
Qy 2079 aagcaagaatgaataaaaaattccaatgtc 2108
|||||
Db 33 AAGCAAGATGAATAAATAATTCCAATGTC 4
|||||

RESULT 8
AW102615/c 473 bp mRNA linear EST 19-OCT-1999
LOCUS
DEFINITION x658h02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2602803 3', mRNA sequence.
ACCESSION AW102615
VERSION AW102615.1 GI:6073228
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 473)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer:-40UP from Gibco
High quality sequence stop: 468.
FEATURES
source
1. 473
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2602803"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not 1; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP-GCBI) were mixed, and ss circles were made as
viro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 156 a 93 c 89 g 135 t
ORIGIN

Query Match 20.8%; Score 443.6; DB 9; Length 473;
Best Local Similarity 98.7%; Pred. No. 4.8e-57;
Matches 468; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Qy 1642 gagcgtctctgcatagaacccaataagttcttctgtttgttttttcaattatttagaaa 1701
|||||
Db 473 GAGGCTGCTGCGATGAACCCCTAATGGTTCCTTGTGTTTTCATAATTATTAGAAA 414
|||||
Qy 1702 taagtctccggatggcgtgtgtgtatcaccacttaaaa-tctctagagaactactgaaca 1760
|||||
Db 413 TAAGTTCTCCGGATGGGCTGTGTGATACCATTAAATTTCTCTAGACAACTACTGAACA 354
|||||
Qy 1761 cctaagaatttctgtagctagattttccccagagacacgcgaactgtcagttttcc 1820
|||||

```

```

Db 353 CCTAA-ATTTTCTAGCGTAGATATTTCCCGAGGACGACGCAACTGTCTAGTCTTTCC 295
Qy 1821 taagcccccgggaagcagcaggaatgggctccagggccaggttcacaccagatgtc 1880
|||||
Db 294 TAAGCCCCCGGGAGACGAGCAATGGGGCTCCGAGGCCAGGCTTGCCACGAGATGC 235
|||||
Qy 1881 ttgagttagaggactaaaaattatccagttcttctgtgttcttctactgaattgtggaaa 1940
|||||
Db 234 TTGAGTTAGAGGACTTAAATATATCCAGTTCTCTGTTCTTACTTGAATGTGGAAA 175
|||||
Qy 1941 agctctatatccaaatgaactctccataattattgttgtaatatattattgtttgtaa 2000
|||||
Db 174 AGCTCTATTATCCCAATTAATCTCTCCATAATATTATTGTTGTAATATTATTATTGTTAA 115
|||||
Qy 2001 aacatggttcacataactagcttctgtggaaacacagcagatgaataatcttaagtac 2060
|||||
Db 114 AACATGGTTACATPACTAGCTTGTGGAAACCGACGAGTAAATGAATCTTAAAGTTGAC 55
|||||
Qy 2061 gcttttggttctgttgaagcaaatgaataaaaaatttccaatttcgaataa 2114
|||||
Db 54 GCTTTTGTCTGTTGTAAGCAAGATGAATAAATAATTTCCAATGCTCTCAAA 1
|||||

RESULT 9
BI792942/c 435 bp mRNA linear EST 01-OCT-2001
LOCUS
DEFINITION ie49e01.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
CDNA 3', mRNA sequence.
ACCESSION BI792942
VERSION BI792942.1 GI:15820667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 435)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
, Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
JOURNAL Endocrine Pancreas Consortium
Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 435.
FEATURES
source
1. 435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site.1: Not 1;
Site.2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from

```


SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 419)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 403.

FEATURES
Location/Qualifiers
1..419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2975881"
/lab_host="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
134 a 83 c 80 g 122 t

BASE COUNT
ORIGIN
Query Match 19.4%; Score 413.8; DB 9; Length 419;
Best Local Similarity 99.5%; Pred. No. 1.4e-52;
Matches 415; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1692 tatttagaaataagttctccggatgggtgtgtgataccacttaaaatctctagagaac 1751
|||||
Db 419 TATTTAGAAATAAGTTCTCCGGATGGGCTGTGTGATACCACCTTAAATCTCTAGAGAAC 360
|||||
Qy 1752 tactgaacacctaagaattttctgtagcgtagatatctcccgagagacacgcgaactgtc 1811
|||||
Db 359 TACTGAACACCTAAAGATTTTCTGTAGGCTAGATATTTCCCGAGGACGCGAACTGTC 300
|||||
Qy 1812 agtcttcttaagcccccgagagacgagcaaatgggctcgcagccaggttcgac 1871
|||||
Db 299 AGTCTTCTTAAGGCCCCCGGAGACGAGCAATGGGGCTTCGAGGCGAGGCTTGAC 240
|||||
Qy 1872 cagcatgtctgagtagaggaacttaaaattatccagttctctgtgttcttacttgaa 1931
|||||
Db 239 CAGCATGCTTGAGTTAGAGGACTTAAATATATCCAGTTCTCTGCTTCTACTTGA 180
|||||
Qy 1932 ttgtggaagactctattatccaataaactctctccatattattgttgtaattattat 1991
|||||
Db 179 TTGTGGAAGAGCTCTATTATCAATTAATTTCTCCATATTTATTTGTAATTTAT 120
|||||
Qy 1992 tttgttgaacatggttcacataactagctgtgtggaacacagcaggttaaatgattct 2051
|||||
Db 119 TGTTTGTAAAACATGTTTACATAACTAGCTGTGTGGAACACGAGGTAATGAATTC 60
|||||
Qy 2052 taagtgcagcttttgggtctgtgtgaaagcaagatgaataaaatttccaatgtc 2108
|||||
Db 59 TAAGTTGACGCTTTTGGTCTGTGTTAAAGCAACATGAATAAATAATTTCCAATGTC 3

RESULT 12
AA846474/c

LOCUS
AA846474 429 bp mRNA linear EST 31-DEC-1998
DEFINITION aj56d11.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394325
3', mRNA sequence.
ACCESSION AA846474
VERSION AA846474.1 GI:2933614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 429)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo,
Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 432 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 413.
Location/Qualifiers
1..429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394325"
/lab_host="Soares_testis_NHT"
/sex="male"

FEATURES
source
1..429

BASE COUNT
ORIGIN
Query Match 19.4%; Score 413.2; DB 9; Length 429;
Best Local Similarity 99.1%; Pred. No. 1.7e-52;
Matches 426; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1682 ttttcaaatatttagaataaagttctccggatgggtgtgtgataccacttaaaatc 1741
|||||
Db 429 TTTTATAAATATTAGAAATAAGTTCTCCGGATGGGCTGTGTGATACCACCTTAAATC 370
|||||
Qy 1742 tctagagaactactgaacacacttaagattttctgtagcgtagatatctcccgagagac 1801
|||||
Db 369 TCTAGAGAACTACTGAACACCTTAAAGATTTTCTGTAGGCTAGATATTTCCCGAGGAC 310
|||||
Qy 1802 gcgaactgtcagttcttctaagcccccgagagacgagcaatgggctcgcagggc 1861
|||||
Db 309 GCGAACTGTCAGTCTTTCTTAGGCCCCCGGGAGACGAGGCAAT- GGGCTCGGAGGCC 251
|||||
Qy 1862 aggttcgaccagcagctgtcttgagtagaggacttaaaattatccagttctctctgtgt 1921
|||||
Db 250 AGGCTTGACACGAGCATGCTTGTAGTTAGAGGACTTAAATATATCCAGTTCTCTGTGT 191
|||||
Qy 1922 tctactgaattgtgaaagactcttatccaataaactctccatattatttggta 1981
|||||
Db 190 TCTACTGAATGTGAAAGCTCTATTATCAATTAATTAATTAATTAATTAATTAATTA 131

Qy	1982	atattattattgttataaacatggttcaataactagcttctggaacaccaggttaa	2041
Db	130	ATATTATTATTGTTGTAAACATGGTTCACATAACTAGCTTTGTGGAACACGACAGTAA	71
Qy	2042	aatgaattcttaagtgcacgttttggttctgttgaagcaagaatgaataaaatttc	2101
Db	70	ATNGAATCTTAGTTCAGCTTTTGGTCTGTTGTTAAACCAAGATGATATAAAATTC	11
Qy	2102	caatgtcgaa	2111
Db	10	CAATGTCAAA	1
RESULT	13		
LOCUS	AI624835/c		
DEFINITION	AI624835 474 bp mRNA linear EST 14-DEC-1999 ts71g06.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2236762 3', mRNA sequence.		
ACCESSION	AI624835		
VERSION	AI624835.1	GI:4649766	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 474)		
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapsb-r@mail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Insert length: 878 Std Error: 0.00 Seq primer: -400P from Gibco High quality sequence stop: 455 POLYA=No.		
FEATURES	Location/Qualifiers		
source	1..474 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2236762" /clone_lib="NCI_CGAP_GC6" /tissue_type="pooled germ cell tumors" /lab_host="DH108" note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	141 a	102 c	92 g 139 t
ORIGIN			
Query Match	19.38;	Score 412;	DB 9; Length 474;
Best Local Similarity	98.88;	Pred. No. 2.5e-52;	
Matches 415; Conservative	0; Mismatches	5; Indels	0; Gaps 0;
Qy	1689	aattatttagaataaagtcttcggatgggctgtgtgataccacttaaaatctctag	1748
Db	424	AATTGTGAAGAAATAAGTTCCCGATGGGCTGTTGTGATACCACTTAAATCTCTAGAG	365

BASE COUNT 147 a 114 c 103 g 145 t

Query Match 19.3%; Score 412; DB 9; Length 509;
Best Local Similarity 98.8%; Pred. No. 2.5e-52;
Matches 415; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1689 aattattagaataagttctcggatggtgtgtgtatccacttaaaatctctagag 1748
DB 423 AATTTGGAGAATAAGTTCTCCGATGGCTGTGTGTATACCACTTAAATCTCTAGAG 364
QY 1749 aactactgaacacccaaagtatttctgtagcgtagatatttccccagagacacgcgaact 1808
DB 363 AACTACTGAACACCTTAAAGATTTCTGTAGCTAGATATTTCCCGAGAGCGCGGAAT 304
QY 1809 gtcagtccttcttaagggccccgggagacgaggaatggggcctcgagggcaggttg 1868
DB 303 GTCAGTCTTTTCCTAAGGCCCGGGAGACGAGGCAATGGGGCTCGCAGGCGAGGCTTG 244
QY 1869 caccagcatgtcttgagttagagacttaaaattatccagtttctctgtgttctactt 1928
DB 243 CACCAGCATGCTCTGAGTTAGAGGACTTAAATATATCCAGTTCTTCTGTGTCTACTT 184
QY 1929 gaattgtgaaagctctattatccaataacttctccataattattgttgaattat 1988
DB 183 GAATTGTGGAAGAGCTCTATTATCCAAATTAACCTCTCCATAATTATTGTGAATATTAT 124
QY 1989 tattgtttgaaacatggttcacataactagcttgtgaaacccagcaggtgaaatgaat 2048
DB 123 TATTGTTTGTAAACATGTTTCCACATAACTAGCTTGTGGAACCCAGCAGGTAAATGAAT 64
QY 2049 tcttaagttgacgctttgttctgtttaaagcaagatgaataaaatttccaatgtc 2108
DB 63 TCTTAAGTTGACGCTTTTGGTTCTGTGTGTAAAGCAAGATGAATAAAATTTCCCAATGTC 4

RESULT 15
LOCUS A1968415/c
DEFINITION wu02h05.x1 NCI_CGAP_GC6 Homo sapiens CDNA clone IMAGE:2515833 3',
mRNA sequence.
ACCESSION A1968415
VERSION A1968415.1 GI:5765233
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbbrp/image/image.html
Insert Length: 1082 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 464.
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2515833"

/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldi.
BASE COUNT 171 a 153 c 141 g 177 t

Query Match 19.3%; Score 412; DB 9; Length 642;
Best Local Similarity 98.8%; Pred. No. 2.3e-52;
Matches 415; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1689 aattattagaataagttctcggatggtgtgtgtatccacttaaaatctctagag 1748
DB 423 AATTTGGAGAATAAGTTCTCCGATGGCTGTGTGTATACCACTTAAATCTCTAGAG 364
QY 1749 aactactgaacacccaaagtatttctgtagcgtagatatttccccagagacacgcgaact 1808
DB 363 AACTACTGAACACCTTAAAGATTTCTGTAGCTAGATATTTCCCGAGAGCGCGGAAT 304
QY 1809 gtcagtccttcttaagggccccgggagacgaggaatggggcctcgagggcaggttg 1868
DB 303 GTCAGTCTTTTCCTAAGGCCCGGGAGACGAGGCAATGGGGCTCGCAGGCGAGGCTTG 244
QY 1869 caccagcatgtcttgagttagagacttaaaattatccagtttctctgtgttctactt 1928
DB 243 CACCAGCATGCTCTGAGTTAGAGGACTTAAATATATCCAGTTCTTCTGTGTCTACTT 184
QY 1929 gaattgtgaaagctctattatccaataacttctccataattattgttgaattat 1988
DB 183 GAATTGTGGAAGAGCTCTATTATCCAAATTAACCTCTCCATAATTATTGTGAATATTAT 124
QY 1989 tattgtttgaaacatggttcacataactagcttgtgaaacccagcaggtgaaatgaat 2048
DB 123 TATTGTTTGTAAACATGTTTCCACATAACTAGCTTGTGGAACCCAGCAGGTAAATGAAT 64
QY 2049 tcttaagttgacgctttgttctgtttaaagcaagatgaataaaatttccaatgtc 2108
DB 63 TCTTAAGTTGACGCTTTTGGTTCTGTGTGTAAAGCAAGATGAATAAAATTTCCCAATGTC 4

Search completed: July 11, 2002, 09:13:05
Job time: 3502 sec

FEATURES
source

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 10:45:37 ; Search time 95.08 seconds
(without alignments)
5513.062 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 2134

Sequence: 1 aattcgggtcaggcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.6	2.3	2028	2	US-09-211-930-12
2	48.6	2.3	2028	3	US-09-340-993-12
3	48.6	2.3	2028	4	US-09-468-442-12
4	48.4	2.3	3234	1	US-08-264-534-31
5	48.4	2.3	3234	1	US-08-083-590A-10
6	48.4	2.3	3234	1	US-08-485-500-31
7	48.4	2.3	3234	2	US-08-346-128-31
8	48.4	2.3	3234	3	US-08-532-384-10
9	48.4	2.3	3234	3	US-08-893-828-31
10	47.4	2.2	1909	3	US-09-100-193-6
11	47.4	2.2	4287	1	US-08-244-189-1
12	47.4	2.2	4287	1	US-08-306-691B-53
13	43.6	2.0	2435	4	US-09-306-593-1
14	42.8	2.0	2447	2	US-09-014-969-14
15	42.8	2.0	2993	2	US-08-415-593-42
16	42.6	2.0	6124	4	US-08-213-419B-3
17	42.4	2.0	6243	2	US-09-056-075-1
18	41.8	2.0	5923	4	US-09-064-922-3
19	41.6	1.9	5467	1	US-07-745-206A-12
20	41.6	1.9	5467	2	US-08-311-363-12
21	41.6	1.9	5852	1	US-07-867-106-2
22	41.4	1.9	739	4	US-08-791-115B-15
23	41.4	1.9	1046	1	US-08-361-467B-4
24	41.4	1.9	1046	1	US-08-484-332C-4
25	41.4	1.9	1117	4	US-09-247-373B-33
26	41.4	1.9	1990	1	US-08-256-130A-1
27	41.4	1.9	3871	2	US-08-599-455B-3

28	41.4	1.9	3971	4	US-09-069-781B-3	Sequence 3, Appl
29	41.2	1.9	1395	2	US-08-553-367A-1	Sequence 1, Appl
30	41.2	1.9	1395	4	US-09-295-306-1	Sequence 1, Appl
31	41	1.9	1493	1	US-08-340-820-24	Sequence 24, Appl
32	41	1.9	1493	1	US-08-593-535-24	Sequence 24, Appl
33	40.8	1.9	573	4	US-08-906-156A-67	Sequence 67, Appl
34	40.6	1.9	1550	2	US-08-609-443B-17	Sequence 17, Appl
35	40.6	1.9	1776	3	US-08-655-352-10	Sequence 10, Appl
36	40.4	1.9	590	4	US-08-906-156A-19	Sequence 19, Appl
37	40.4	1.9	972	5	PCT-US95-04801-2	Sequence 2, Appl
38	40.4	1.9	1101	3	US-08-987-904A-1	Sequence 1, Appl
39	40.2	1.9	1804	2	US-08-504-459-5	Sequence 5, Appl
40	40.2	1.9	6232	4	US-08-456-200B-11	Sequence 11, Appl
41	40	1.9	2065	4	US-09-370-473-5	Sequence 5, Appl
42	40	1.9	6854	4	US-09-194-905-7	Sequence 7, Appl
43	40	1.9	19124	2	US-08-487-826B-13	Sequence 13, Appl
44	39.6	1.9	3200	1	US-08-444-405-1	Sequence 1, Appl
45	39.6	1.9	3200	1	US-08-384-850-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-211-930-12
; Sequence 12, Application US/09211930
; Patent No. 5962265
; GENERAL INFORMATION:
; APPLICANT: Tyrell E. No. 5962265rls
; APPLICANT: William Craig Moore
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION
; TITLE OF INVENTION: SERINE/THREONINE KINASE
; FILE REFERENCE: PHM.70296
; CURRENT APPLICATION NUMBER: US/09/211.930
; EARLIER FILING DATE: 1998-12-15
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-211-930-12

Query Match	2.3%	Score 48.6;	DB 2;	Length 2028;
Best Local Similarity	51.3%;	Pred. No. 0.007;		
Matches 138;	Conservative 0;	Mismatches 129;	Indels 2;	Gaps 1;
Qy	1868	gcacacagatgcttgaggttagaggaactaaattatccagttttctctgtgttctact	1927	
Db	1738	gtacattattcttctgttcattgatgaatctgtttggttggaactcttagttaac	1797	
Qy	1928	tgaatttggaagactcta--ttatccaattaaactctccataattattgttgaatat	1985	
Db	1798	tgtatactcaaacatacaggtctttccaaagtcattcttaactataaattgttgaatat	1857	
Qy	1986	tattattgttgtaaacattggttcacataactagcttgttggaacacagcaggtaaatg	2045	
Db	1858	catcaagcttcaaaaagcattctttcccccacacagtagtatctctaaatgactatt	1917	
Qy	2046	aattcttaagtgcgcgttttgggttctgttgaagcaaaagatgaataaattcccaat	2105	
Db	1918	tgtaatagggtgggaagtaagtaatactcttcttaaacacagtggtttttaagaagctccgg	1977	
Qy	2106	gtcgaataaaaaaataaaaaaataaaaaa 2134		
Db	1978	aaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2006		
RESULT	2			

```

US-09-340-993-12
; Sequence 12, Application US/09340993
; Patent No. 6034228
; GENERAL INFORMATION:
; APPLICANT: Tyrell E. No. 6034228r1s
; APPLICANT: William Craig Moore
; APPLICANT: David Shay Silberstein
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE
; FILE REFERENCE: PHM.70296.N1
; CURRENT APPLICATION NUMBER: US/09/340.993
; CURRENT FILING DATE: 1999-06-25
; EARLIER APPLICATION NUMBER: GB 9726851.0 & US 09/211,930
; EARLIER FILING DATE: 1997-12-19 & 1998-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-340-993-12

```

	Query Match	2.38;	Score 48.6;	DB 3;	Length 2028;
	Best Local Similarity	51.3;	Pred. No. 0.007;		
	Matches 138;	Conservative 0;	Mismatches 129;	Indels 2;	Gaps 1;
Qy	1868	gcaccagcatgcttggttagtgaggacttaaaattatccagtttcttctgtgttctact	1927		
Db	1738	gtaccattattcttaigtctcatggaaagataactgtttggtttggaaatctttagttaac	1797		
Qy	1928	tgaatttggaagactcta--ttatccaattaaactttctccataattattgttgtaaat	1985		
Db	1798	tgtatactcataacatacaggtgtttccaagtcatacttaactataatgtttgtaaat	1857		
Qy	1986	tattattgtttgtaaaacatggttcacataactagcttgtggaaaccagcagggtaaaa	2045		
Db	1858	catcaagcttcaaaaagcattctttttcccccacacaagtatttctcaaaatgactatt	1917		
Qy	2046	aattcttaagttgacgcttttgggtctgtttgtaagcaagatgaataaaatttccaat	2105		
Db	1918	tgtaatgaggcggaagtgaagcaatccttctcaaaacaaagtgtttttgaagaagctccogg	1777		
Qy	2106	gtcgaaaaaataaaaaaataaaaaa	2134		
Db	1978	aaaaaataaaaaaataaaaaaataaaaaa	2006		

```

RESULT      3
US-09-468-442-12
; Sequence 12, Application US/09468442
; Patent No. 6300098
; GENERAL INFORMATION:
; APPLICANT: Tyrell E. No. 6300098:ris
; APPLICANT: William Craig Moore
; APPLICANT: David Shay Silberstein
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION SERINE/THREONINE KINASE
; FILE REFERENCE: PHM 70296.N1
; CURRENT APPLICATION NUMBER: US/09/468,442
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/340,993
; EARLIER FILING DATE: 1999-06-25
; EARLIER APPLICATION NUMBER: GB 9726851.0
; EARLIER FILING DATE: 1997-12-19
; EARLIER APPLICATION NUMBER: US 09/211,930
; EARLIER FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-468-442-12

```

Query Match	2.3%;	Score 48.6;	DB 4;	Length 2028;
Best Local Similarity	51.3%;	Pred. No. 0.007;		
Matches 138;	Conservative	0;	Mismatches 129;	Indels 2;
Gaps				
Qy 1868	gcaccagcatgtcttgagttagaggactiaaaattccoaagttctctgtgtttctact	1927		
Db 1738	gtaccattattcttatgtctcaggaatgaactgttggtttggaaatcttttagttaac	1797		
Qy 1928	tgaattgtggaaaagctcta--ttatccaattaaacttctccaattattgttgttaatat	1985		
Db 1798	tgtatactataacatacacagggtctttccaaagtcactcctaactattaaagtgttgaatat	1957		
Qy 1986	tattattgtttgtaaaacatggtttcacataactagtgttggaaaacagcaggtataaatg	2045		
Db 1858	catcaagcttcaaaaagcattctttttccccccacacagtatattcttaaaaatgactatt	1917		
Qy 2046	aattcttaattgacgctttgtgttggtttgaagcaagatgaaataaaaatttccaat	2105		
Db 1918	tgtaattgagtggaagtaagataacacctctttaaacaacaaagtgttttaagaagctcccg	1977		
Qy 2106	gtcgaaaaaataaaaaaataaaaaaataaaaaa 2134			
Db 1978	aaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2006			

```

RESULT      4
US-08-264-534-31
; Sequence 31, Application US/08264534
; Patent No. 5648464
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakanos, Spyridon et al.
; TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,534
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/695,189
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3234
US-08-264-534-31

```


;; TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON
;; NUMBER OF SEQUENCES: 34
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: PENNIE & EDMONDS
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10036-2711
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/893,828
;; FILING DATE: 11-JUL-1997
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mistock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 7326-050
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-8864/9741
;; INFORMATION FOR SEQ ID NO: 31:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3234 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: cDNA
US-08-893-828-31

Query Match 2.3%; Score 48.4; DB 3; Length 3234;
Best Local Similarity 55.3%; Pred. No. 0.01;
Matches 94; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 18 aagctcgggggggacagcagcgagcggtctctcggggagtagtaccgggagatccca 77
Db 289 AACAGCGGGAGTGGAGTGGGGGCTGGACTGTGCGGAGCATGTACCCGAGAGGCTG 348
Qy 78 gagctcagcgctggagagcgctcctcagcgcccgagcgcccgctgcaac 137
Db 349 GCGCGGCACGCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 408
Qy 138 gtgactctctctctctctctctctctctctctctctctctctctctctctctctct 187
Db 409 TTCACCTTCTTCGCGGAGCTCAGCGCGCTGCTGCACACACCAACGTGGTCTT 458

RESULT 10
US-09-100-193-6
; Sequence 6, Application US/09100193
; Patent No. 6153729
; GENERAL INFORMATION:
; APPLICANT: Gary S. Stein et al.
; TITLE OF INVENTION: NUCLEAR MATRIX TARGETING PEPTIDES AND USES THEREFORE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/100,193
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/050,104
;; FILING DATE: 20-JUNE-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jane E. Remillard
;; REGISTRATION NUMBER: 38,872
;; REFERENCE/DOCKET NUMBER: UMM-024
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 227-7400
;; TELEFAX: (617) 742-4214
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1909 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 766..1515
US-09-100-193-6

Query Match 2.2%; Score 47.4; DB 3; Length 1909;
Best Local Similarity 55.0%; Pred. No. 0.014;
Matches 93; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 8 ggtcaagcggaagctcgcggggcgacgacgagcggggagctcctcgcgggagtagtacc 67
Db 837 GATGAGCGAGGCGTGGCGTGGGGCGCCCGGACGCGCGCTGCGCTGCGCTGCGCTG 896
Qy 68 cggatccagagctcagcgctgagcgctgagcgctcctcagcgcccgagcgcccgcc 127
Db 897 GAGGAGCGGCGACGCGACGATGGTGGAGGTGCTGGCCGACACCGCGGCGAGTGGT 956
Qy 128 cgcctgaacgtgcacttcctgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 176
Db 957 CACCGACAGCCCCAACTTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005

RESULT 11
US-08-244-189-1
; Sequence 1, Application US/08244189
; Patent No. 5580727
; GENERAL INFORMATION:
; APPLICANT: Ohki, Misao
; APPLICANT: Kikuchi, Kimiko
; APPLICANT: Miyoshi, Hiroyuki
; APPLICANT: Kozu, Tomoko
; TITLE OF INVENTION: No. 5580727el DNAs, Polypeptides Encoded Thereby
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,189
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.

QY 2015 aactagcttggaaccagcggttaattcttaagttaagctgttttggttctgt 2074
Db 2107 CACTTACTATAGCAAAAGAGTAAATAATCCATTATTAATATATTTATATAA 2048
QY 2075 tctaaagcaagatgaataaaatttccaa 2104
Db 2047 TTGTAATAAATATTATTATATAATTTTAA 2018

RESULT 14

US-09-014-969-14

; Sequence 14, Application US/09014969

; Patent No. 5965397

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth

; APPLICANT: McCoy, John M.

; APPLICANT: Lavallie, Edward R.

; APPLICANT: Racie, Lisa A.

; APPLICANT: Merberg, David

; APPLICANT: Treacy, Maurice

; APPLICANT: Spaulding, Vikki

; APPLICANT: Agostino, Michael J.

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: MA

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/014,969

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sprunger, Suzanne A.

; REGISTRATION NUMBER: 41,323

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 498-8284

; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2447 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

; US-09-014-969-14

Query Match 2.0%; Score 42.8; DB 2; Length 2447;
Best Local Similarity 58.3%; Pred. NO. 0.24;
Matches 91; Conservative 1; Mismatches 63; Indels 1; Gaps 1;

QY 1980 taatattattattgttgaacatggttcacataactagct-tgtggaaccagcagg 2038
Db 2152 TAATGGCTGATAGATATGATATATGGTTACATATCTGGATCTGTGTTGTTG 2211
QY 2039 taaatgaattcttaagttgacgcttttggttctgttgaagcaagaatgaataaaat 2098
Db 2212 TACTTTAAATGTGACAAATAAACCTTTGGGAGAAAAAATAAAAAA 2271

QY 2099 ttccaatgtcgaataaaataaaataaaataaaataaa 2134

Db 2272 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2307

RESULT 15

US-08-415-593-42

; Sequence 42, Application US/08415593

; Patent No. 5912140

; Patent No. 5912140 5776726

; GENERAL INFORMATION:

; APPLICANT: Whoriskey, Susan K.

; APPLICANT: Quinn, Cheryl L.

; APPLICANT: Tao, Nisjun

; APPLICANT: Politis-Virk, Karen I.

; APPLICANT: Schimmel, Paul R.

; TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA

; TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: Massachusetts

; COUNTRY: U.S.

; ZIP: 02173

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/415,593

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Brook, David E.

; REGISTRATION NUMBER: 22,592

; REFERENCE/DOCKET NUMBER: CFI94-09

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-861-6240

; TELEFAX: 617-862-9540

; INFORMATION FOR SEQ ID NO: 42:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2993 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..2934

; US-08-415-593-42

Query Match 2.0%; Score 42.8; DB 2; Length 2993;
Best Local Similarity 51.6%; Pred. No. 0.26;
Matches 98; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1945 ctattatccaatttaacttctccataattattgttggaatattattattttgttaaaaca 2004
Db 2803 CTTTAAACCAAAATCCTTCGTAGGCCTTTAGAAAAAATCAATCAATATATTAGATGAACA 2862
QY 2005 tgggttcacataactagcttggtaaacaccagcaggtaaaatgaattcttgaattgcgctt 2064
Db 2863 GATCCTAAACAGATTGTAAGAAAGAAAAACAGGATGTTCAAGGCGCAACATTTATGTATCT 2922
QY 2065 tgggttcgttgaagcaagcaagataaaattcccaatgcgaataaaataaaataaa 2124
Db 2923 TTATTACGATTATAAATATTATTATAAATAAATCAATTAATATTTCAAAAAA 2982
QY 2125 aaaaaaaa 2134
Db 2983 AAAAAAAAAA 2992

Fri Jul 12 08:35:25 2002

Search completed: July 11, 2002, 10:46:24
Job time: 9101 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 12:28:35 ; Search time 5652.01 Seconds
(without alignments)
8168.576 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aattcggggtcaaggcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 21979536 seqs, 1081749327 residues 43959072
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2134	100.0	2134	29	US-09-743-237-3
2	2101	98.5	2638	25	US-09-643-893-287
3	1631.6	76.5	1770	18	US-09-471-275-5610
4	662	31.0	1861	29	US-09-743-237-2
5	647.8	30.4	2241	29	US-09-743-237-1
6	633.8	29.7	2201	75	US-60-360-207-983
7	476.8	22.3	3495	23	US-09-611-526-2894
8	429	20.1	445	1	PCT-US01-01329-425
9	429	20.1	445	1	PCT-US01-01339-1301
10	429	20.1	445	30	US-09-764-891-1301
11	429	20.1	445	37	US-10-080-090-425
12	404	18.9	465	17	US-09-362-510-45961
13	404	18.9	465	17	US-09-362-510A-45961
14	404	18.9	465	34	US-09-904-013-45961
15	401.4	18.8	431	13	US-08-971-197-11928
16	401.4	18.8	431	13	US-08-971-197C-11928
17	394.4	18.5	397	16	US-09-289-768-13239
18	394.4	18.5	397	35	US-09-939-397-13239
19	308	14.4	441	18	US-09-489-036-32385
20	308	14.4	441	35	US-09-943-143-32385
21	270.8	12.7	302	14	US-09-013-812-1497
22	270.8	12.7	302	21	US-09-540-764-42635
23	270.8	12.7	302	42	US-60-036-570-1497
24	249	11.7	249	20	US-09-539-806-21625
25	245	11.5	250	1	PCT-US00-26524B-2645
26	230	10.8	444	18	US-09-404-549-314
27	230	10.8	444	18	US-09-404-549A-314
28	221.6	10.4	4276	75	US-60-360-207-19806
29	203.8	9.6	3535	14	US-09-087-136-16
30	203.8	9.6	3535	14	US-09-087-136-16
31	203.8	9.6	3535	16	US-09-220-091-16


```

1441 actcagagctctctgaagatgtggcaaccatgccctcttcttgagagtgcatggcc 1500
1501 tgagcattgtttgtgtggtccagagagagagagctgttggttcccatagctcctgggagagtg 1560
1501 tgagcattgtttgtgtggtccagagagagagagctgttggttcccatagctcctgggagagtg 1560
1561 tctcagagcgagagagagagagagagagagagagagagagagagagagagagagagagagagag 1620
1561 tctcagagcgagagagagagagagagagagagagagagagagagagagagagagagagagagag 1620
1621 ctctcagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1680
1621 ctctcagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1680
1681 gtttttcaaatattattagaaataagttctcggagagagagagagagagagagagagagagagag 1740
1681 gtttttcaaatattattagaaataagttctcggagagagagagagagagagagagagagagagag 1740
1741 ctctagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1800
1741 ctctagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1800
1801 cgcgaactgtcagttcttcttaagagagagagagagagagagagagagagagagagagagagag 1860
1801 cgcgaactgtcagttcttcttaagagagagagagagagagagagagagagagagagagagagag 1860
1861 caggcttcacacagagagagagagagagagagagagagagagagagagagagagagagagagag 1920
1861 caggcttcacacagagagagagagagagagagagagagagagagagagagagagagagagagag 1920
1921 tctcactggaattgtggaaagagagagagagagagagagagagagagagagagagagagagagag 1980
1921 tctcactggaattgtggaaagagagagagagagagagagagagagagagagagagagagagagag 1980
1981 aatattattattgtttgaaacatagtttcacataactagcttcttggaacacagagagagagag 2040
1981 aatattattattgtttgaaacatagtttcacataactagcttcttggaacacagagagagagag 2040
2041 aatgaattattattgtgagagagagagagagagagagagagagagagagagagagagagagagag 2100
2041 aatgaattattattgtgagagagagagagagagagagagagagagagagagagagagagagagag 2100
2101 ccaatgtcgaagagagagagagagagagagagagagagagagagagagagagagagagagagag 2134
2101 ccaatgtcgaagagagagagagagagagagagagagagagagagagagagagagagagagagag 2134

```

RESULT 2

```

US-09-643-893-287
; Sequence 287, Application US/09643893
; GENERAL INFORMATION:
; APPLICANT: Silos-Santiago, Inmaculada
; APPLICANT: DiStefano, Peter
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1161-001
; CURRENT APPLICATION NUMBER: US/09/643.893
; CURRENT FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: 60/150,147
; PRIOR FILING DATE: 1999-08-20
; NUMBER OF SEQ ID NOS: 412
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 287
; LENGTH: 2638
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2638)
; OTHER INFORMATION: n = A,T,C or G
US-09-643-893-287

```

```

Query Match 98.5%; Score 2101; DB 25; Length 2638;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2104; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 cggggtcaagcgagagagagagagagagagagagagagagagagagagagagagagagagagagag 64
Db 260 cggggtcaagcgagagagagagagagagagagagagagagagagagagagagagagagagagagag 319
Qy 65 cccggggtccagagagagagagagagagagagagagagagagagagagagagagagagagagagag 124
Db 320 cccggggtccagagagagagagagagagagagagagagagagagagagagagagagagagagagag 379
Qy 125 gccggggtccagagagagagagagagagagagagagagagagagagagagagagagagagagagag 184
Db 380 gccggggtccagagagagagagagagagagagagagagagagagagagagagagagagagagagag 439
Qy 185 gttttgccccctggggcgctggttctcgaagagagagagagagagagagagagagagagagagagag 244
Db 440 gttttgccccctggggcgctggttctcgaagagagagagagagagagagagagagagagagagagag 499
Qy 245 tcccagttgaaatcaagagagagagagagagagagagagagagagagagagagagagagagagagagag 304
Db 500 tcccagttgaaatcaagagagagagagagagagagagagagagagagagagagagagagagagagagag 559
Qy 305 actttgcaagaatcttctgctcaggaatctctgttgaagagagagagagagagagagagagagagagag 364
Db 560 actttgcaagaatcttctgctcaggaatctctgttgaagagagagagagagagagagagagagagagag 619
Qy 365 gaggatgctcctgctgttcttcttaagagagagagagagagagagagagagagagagagagagagag 424
Db 620 gaggatgctcctgctgttcttcttaagagagagagagagagagagagagagagagagagagagagag 679
Qy 425 aaaggggacacacaaatgctatgtatagacaaatcttagaagaagagagagagagagagagagagagag 484
Db 680 aaaggggacacacaaatgctatgtatagacaaatcttagaagaagagagagagagagagagagagagag 739
Qy 485 catttggttctcagatcaagagatcaaaataatctacagagagagagagagagagagagagagagagag 544
Db 740 catttggttctcagatcaagagatcaaaataatctacagagagagagagagagagagagagagagagag 799
Qy 545 atgactgttttagtagggagagatcttctgagagagagagagagagagagagagagagagagagagagag 604
Db 800 atgactgttttagtagggagagatcttctgagagagagagagagagagagagagagagagagagagagag 859
Qy 605 caacttgaggagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 664
Db 860 caacttgaggagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 919
Qy 665 ctccagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 724
Db 920 ctccagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 979
Qy 725 ttttgcaacaactgcaattgttaataatgttgcaacaactgcaatgcatgatatgaaagagagagag 784
Db 980 ttttgcaacaactgcaattgttaataatgttgcaacaactgcaatgcatgatatgaaagagagagag 1039
Qy 785 ttttaagccattgaagagagagagagagagagagagagagagagagagagagagagagagagagagag 844
Db 1040 ttttaagccattgaagagagagagagagagagagagagagagagagagagagagagagagagagagag 1099
Qy 845 aaggggcaattgggcaatgtcaagcccgagagagagagagagagagagagagagagagagagagagag 904
Db 1100 aaggggcaattgggcaatgtcaagcccgagagagagagagagagagagagagagagagagagagagag 1159
Qy 905 ggctgcttgaagaattactgagagagagagagagagagagagagagagagagagagagagagagagag 964
Db 1160 ggctgcttgaagaattactgagagagagagagagagagagagagagagagagagagagagagagagag 1219
Qy 965 aaatgcatgtgtgcaaaaattatgaagaagagagagagagagagagagagagagagagagagagagag 1024
Db 1220 aaatgcatgtgtgcaaaaattatgaagaagagagagagagagagagagagagagagagagagagagag 1279

```


Db	1590	TTAAATCTCATTTACACAACAACCTTGAGGGAGCGCTTACCATTCCGTTAGTCAACGGGCTCTGCT	1531
Qy	647	tccccctcgggatcaactctccaggaccaccacaaaaataactcttgctcgggctactgtgac	706
Db	1530	TTCCCTTCGGATCACTCTTCCAGGACCACCNAAAATACTTTGGCTGGGTACTGTGAC	1471
Qy	707	tgtttgccagtggggaacttttgcacaactgcgaattgtaataattgttgcacaacttg	766
Db	1470	TGCTTTTGCAGATGGGACCTTTTGCACAACATGCAATTTCTAATAATTGTGTGCAACAACCTTG	1411
Qy	767	catcatgatattgaacgggtttaaagccattaaaggcatgtcttgtagaanaatccagaagct	826
Db	1410	CATCATGATATTGAACGGTTTAAAGCCATTAAAGGCATGCTTGTGTGAAATATCCAGAAGCT	1351
Qy	827	ttccagccaaaaatttggaaggccaaattgggcaatgtccaagccccagcacacaagaagg	886
Db	1350	TTCCAGCCAAAATTTGGGAAGGCCCAATTTGGCAATGTCAAGCCCCAGCACACAAGGG	1291
Qy	887	tgcactgcaggaggttcaggctgcctgaagaattactcgagtgctatgagggccaaatt	946
Db	1290	TGCAACTGCAGAGGTCAAGCTGCCTGAAGAATTACTCGGAGTGTATGAGGCCCAATTT	1231
Qy	947	atgtgtctctatttgcaaatgcattggttgcaaaattatgaagaaagcccaagaacga	1006
Db	1230	ATGTGTTCTCTATTGCAAAATGCAATTTGGTTGCAAAAATATGAAGAAAGCCCCAGAACA	1171
Qy	1007	aegacactaatgagcatgcacaactcacatgcagactggaggttttggaaaggcagcattac	1066
Db	1170	AAGACACTAATGAGCATGCCAACTTACATGCAGACTGGAGGTTTGGAAAGCGCGCATTAC	1111
Qy	1067	ctgccaccacaagaaattttcaggaaattccaagattcagtcacgatagcggcctctctca	1126
Db	1110	CTGCCACCAACGAAATTTTCAGGACTTCCAAGATTCACTACGATAGCGGCGCTTCCTCA	1051
Qy	1127	tgcattcctctggagggttgtagagccacatgcgcctgcctgcttgcctcaggggagaag	1186
Db	1050	TGCATCTCTCTGGAGGTGTGTGAGGCCACATCGCGCTGCCCTGTCTAGGGGAAGAG	991
Qy	1187	gccgaaagaagaaactgctcccaagtgccttggcagagcagaTgaTccttggaggaaatttgg	1246
Db	990	GCCGAAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCTCTGGAGGAATTTGGA	931
Qy	1247	agggtcttatcacagattctccacactgagtttaaatctaaggattgaaaaatggagt	1306
Db	930	AGGTGCTTTATCAGATTTCTCCACACTGAGTTTAAATCTAAGGGATTGAAAATGGAGTAG	871
Qy	1307	agtataaagtgtgaatgcatgttatttgccttagtctagataaatactctagttagaag	1366
Db	870	AGTATAAAGTGTGAATGCATGTTGATTTTGTCTTAACTAGAAATCTCTAGTTTAGAAG	811
Qy	1367	gattgttaggggaacatgagctggctctctgcagcaacaacaggctccccctgcatccctg	1426
Db	810	GATGTTTAGGGAAACATGAGGCTGGCTCTGCAGCAACAACACAGGCTCCCCTGATCCCTG	751
Qy	1427	ggccccggagtttactcagagctctctgaagatgtggcaaccocatgcccccttttctga	1486
Db	750	GGCCCCAGGAGTTTACTCAGAGCTCTCTGAAGATGTGGCAACCCATGCCCTTTTCTGA	691
Qy	1487	ggaggTgcattggcctgagcattgtttgtctggccagaggagagacttgggttccccata	1546
Db	690	GGAGGTGCATGGCCCTGAGCATTTGTTGTCTGGCCCAGAGGAGAGCTTGGGTTTCCCAT	631
Qy	1547	gtccttgggagagtctctgcaggcggc-----	1573
Db	630	GTCTTGGGAGAGTGTCTGCAGGGCGCGGGAGGGCAGACAGGGGAGGGGAGGCACAGCA	571
Qy	1574	-----ggagggcagagcaggccctgcggagagctcac	1605
Db	570	GGCCCTTCGCAAGGGCAGAGCGGGCAGGGGAGGCAGAGCAGGCCCTTGGCGAGAGCTCAC	511
Qy	1606	tctggtcgactcttctctcagagaattgtctctgaggctgtctctgtcatTgaaaacccct	1665
Db	510	TCGTGTCGACTCTTCTCTCAGAGAATGTGCTCTGAGGCTGTCTCTGCATGAAAACCTT	451

Qy	1666	aatggtttcttctgtttgttttcaaatatttagaataaagttctcccgatggggtgttgt	1725
Db	450	AATGGTTCTCTGTTGTTTTCAAATTAATTAGAAATAAGTTCTCCGATGGCTGTTGT	391
Qy	1726	gataceacttaaaatcctctagagaactactgaacacctaagagattttctgtagcgtagat	1785
Db	390	GATACCACCTTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTTCTGTAGCGTAGAT	331
Qy	1786	atttccccagagacacgcgaactgtcagttcttcttaagcccccgagagacgcaggcaa	1845
Db	330	ATTTCCCCACAGGACGCGAACTGTCAAGCTTTCTCTAAGGCCCCGGGACGACGAGCAA	271
Qy	1846	tggggccctcagcaggcagcctgcaccacacatgctcttgagttagaggacttaaaattatc	1905
Db	270	TGGGGCTCTCAGCGCCAGCGTTGCACACGACATGCTCTGAGTTAGAGGACTTTAAATATTC	211
Qy	1906	cagttctctctgttcttacttctgaattgtgaaaagctcttatccaataaactcttc	1965
Db	210	CAGTTCTCTCTGTTTCTTACTTGAATTGTGGAAAAGCTCTATTATCCAAATACTTCTC	151
Qy	1966	cataattatgttgtaattattattgtttgtaaaacatggttcacataactagcttgt	2025
Db	150	CATAATTAATTGTTGTAATATTATTATTGTTTGTAAACATGGTTCACATAACTAGCTTGT	91
Qy	2026	gaaaaccagcagtaaaatgaattcttaagttgacgccttttggtctgttctaagcaca	2085
Db	90	GGAANCCAGCAGGTAAGTAATCTTAAAGTTGACGCTTTTGGTTCTGTGTAAGGCAA	31
Qy	2086	gatgaataaaaatttccaatgcgaaaaa	2115
Db	30	GATCAATAAAAAATTTCCAATGTCTTCAAAA	1
RESULT 4			
US-09-743-237-2			
; Sequence 2, Application US/09743237			
; GENERAL INFORMATION:			
; APPLICANT: SUGIHARA, TAKASHI			
; APPLICANT: WADHWA, RENU			
; APPLICANT: KAUL, SUNIL C.			
; APPLICANT: MITSUI, YOUNG			
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR			
; FILE REFERENCE: 084335/0127			
; CURRENT APPLICATION NUMBER: US/09/743,237			
; CURRENT FILING DATE: 2001-06-04			
; PRIOR APPLICATION NUMBER: PCT/JP99/03859			
; PRIOR FILING DATE: 1999-07-16			
; NUMBER OF SEQ ID NOS: 24			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 2			
; LENGTH: 1861			
; TYPE: DNA			
; ORGANISM: Mus musculus			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (271)..(1155)			
US-09-743-237-2			

Query Match	31.0%	Score 662;	DB 29;	Length 1861;
Best Local Similarity	74.7%	Pred. No. 1.1e-96;		
Matches 863;	Conservative	0;	Mismatches 280;	Indels 13; Gaps 2;

Qy	183	gttttttgccttcgggcgcctgggtcctctgcgaagagcctccaccgcgggcgtccgcac	242
Db	48	gtcgttacttccccggtgcgcggccgcgcgtgtgcgcctaagcgcgaggaagtgcgcgt	107
Qy	243	gattccagttgaaatacaagtagcaggggtgtactactacaagtaataatccgcgaagaag	302
Db	108	gattccagttgaaatacaaaag-aagcaggggtgtagtgtgcagggcgagccctgaagaag	166
Qy	303	caactttgcagaatacttctgtccagaataacctgttgcgaagttcccatagtcccaagaac	362

```

Db 167 cagcttcaggccctctggtcaggaaactctgttgaagttcccatccaggagg 226
QY 363 tagagatgctcctctgtcttcttaagaagaatcccaacaaatggtgatgccaat 422
Db 227 cagagaggctccagctcctcggaaagaagactccagcccccattggtatgtcagc 286
QY 423 tgaaggggggcacacaaatctatgtatagacaaatctagaagaagaaactaaagcac 482
Db 287 tgaaggaggcccgagatgctctgcatagacaaactgtgagggagctcgaagcgc 346
QY 483 tccatttggcttcagtgatcaagatcaaaataattatctacagtcagatgtccctaaca 542
Db 347 tccatttggcttcagtgatcaagatcaaaataattatctacagtcagatgtccctaagc 406
QY 543 caatgactcttagtaggagattttccagcatcaaaataattatctacattacac 602
Db 407 caatgacacttttagtgggaagactctccagtlaccaggaagttaattctcatcacac 466
QY 603 acaacttgaggagcttaccatcggttagtcaacgggtctgttccctcgggataca 662
Db 467 aggttataatgagcttcccatcagctgcaatgggctgcttccctcctggacctg 526
QY 663 ctcttcaggaccacaaataactttgggtggttactgtgactgttgcagtgagg 722
Db 527 ctctgcaaggccacccaaataactctgtctggttactgtgactgttctccagcggg 586
QY 723 acttttgcaacactcaattgaataattgttgcaacactgtcatatgatattgaac 782
Db 587 actctgcaacactgca-----gctgcaacactgcccagctgagctgagc 634
QY 783 ggttaagcatttaaggatgctctgttggtagaataccagaagcttccagccaaattg 842
Db 635 gcttcaagcctaagaaggctctgtgtagaatactgaagcttccacaaataagg 694
QY 843 ggaaggccaattggccaatgtcaagcccccagcaacaaagggtgcaactgcaggagt 902
Db 695 ggaaggccgctggagctgtcaactctgcacagcaaaagggtgcaactgtaagcgt 754
QY 903 caggctgctgaagaacttactgcagtgctatgagcccaaatatgtgttctctatt 962
Db 755 caggctgctgaagaacttactgcagtgctatgagcccaaatatgtgttctctatt 814
QY 963 gcaaatgcttgggtgcaaaattatgaagaagccagaaacacacactaatgagca 1022
Db 815 gcaaatgcttgggtgcaaaattatgaagaagccagaaacacacactaatgagca 874
QY 1023 tgccaaactacatgcagactgagggtttggaaggcagccattacctgccacacgaat 1082
Db 875 caccctactacatggagcctgggactttgagagcagccaattattgtcccccagcaat 934
QY 1083 ttccaggacttccaagattcagtcagatagcgccgtcttctctcatgcatctcctggagg 1142
Db 935 tctcaggacttccaagactgagaaaaataaggcagcccttctctctgtatctcctggagg 994
QY 1143 tggaggagccacatgcgctgctgtctcagggagagagggccgagaaagaaact 1202
Db 995 tagtggaggccaatgctgctgctgctgcccaggggtgaggaagcagagcaggagcaat 1054
QY 1203 gctccagtgcttcgagacagatgactcctggaggaaatttgaagggtgcttatcaga 1262
Db 1055 gttcccaagcttggtgagcagatgactcctggaggaggttgaagggtgcttcgaga 1114
QY 1263 ttctccactagtttcaaatcctaaaggattgaaatggagtagagataaagtgtgaa 1322
Db 1115 ttctccactagtttcaaatcctaaaggattgaaatggagtagagataaagtgtgaa 1174
QY 1323 gcatgttgatttgc 1338
Db 1175 ggggaatgctgtggc 1190

```

RESULT 5

```

US-09-743-237-1
; Sequence 1, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YUJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (651)..(1535)
US-09-743-237-1

```

Query Match 30.4%; Score 647.8; DB 29; Length 2241;
 Best Local Similarity 76.1%; Pred. No. 2.1e-94;
 Matches 829; Conservative 0; Mismatches 247; Indels 13; Gaps 2;

```

QY 250 gttgaatacaggttaagcaggtggtactactacaagtaataatccggaagaagcaacttt 309
Db 495 gttgaatacaggttaagcaggtggtactactacaagtaataatccggaagaagcaacttt 553
QY 310 gcagaatcttctgtcaggaatcctgttgcaggttcccatgttcccatgtccaggaactagaga 369
Db 554 ccaggccctctggtcaggaatcctgttgcaggttcccatgttcccatgtccaggaactagaga 613
QY 370 tgcctcctgctgtctcttaagaagaatcccaacaaatgggtgatgccaattgaaagg 429
Db 614 ggctccagctgctcctcggaaagaagactccagcccatgggtgattgtcagctgaaagg 673
QY 430 gggcacacaaatgctatgtatagacaattctagaacaagaagaaactaaagcactcattt 489
Db 674 agggcccgagatgctctgcatagacaactgtgagcagggagagctcaagcgtccatct 733
QY 490 ggttcctcagtatcaagatcaaaataattatctacagtcagatgtccctaaacacatgac 549
Db 734 gcttcctcagtagatgacagagcaggtttccctcagtcagagctccctaaagcgaatgac 793
QY 550 tgccttagtagggagattttgcccagcatcaacaaataataatctcatcacacaact 609
Db 794 aactttagtgggaagactctcgcagtcagcgaagttaataatctcatcacacaggttga 853
QY 610 tgaggagccttaccatcagtagcaacgggtctgttccctcgggacacactcttcc 669
Db 854 taatgagctctcccatcagctgcaatggggtgcttccctcgggacactgctgca 913
QY 670 aggaccacacaaataaactttgggtggtgactgtgactgttgcagtgaggagactttg 729
Db 914 agggccacacaaataaactctggtggtgactgtgactgttgcagtcagcgggacttctg 973
QY 730 caacacactgcaattgtaataattgttgcaacaaactgcatcatgatatgaaagggttaa 789
Db 974 caacagctgca-----gctgcaacacactgcccagctgagctgagcgtccaa 1021
QY 790 agccattgaagcagctctgttggtagaataccagagctttccagcccaaaaaattgggaagg 849
Db 1022 agccataaaggcgtgcttgatagaaatcctgaagctttccacacaaaaattgggaagg 1081
QY 850 ccaattgggcaatgtcaagccccagcacaacaaagggtgcaactgcaggaggcagctg 909
Db 1082 ccgtctgggagctgctaaacttcgacagcaaaagggtgcaactgtaagcgtcagcgtg 1141

```


; APPLICANT: KOGA, HISASHI
 ; TITLE OF INVENTION: PRIMERS FOR SYNTHESIS OF FULL LENGTH cDNAs
 ; FILE REFERENCE: 08335/0122
 ; CURRENT APPLICATION NUMBER: US/09/611,526
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: JP 1999-194486
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: JP 2000-118774
 ; PRIOR FILING DATE: 2000-01-11
 ; PRIOR APPLICATION NUMBER: JP 2000-183765
 ; PRIOR FILING DATE: 2000-05-02
 ; NUMBER OF SEQ ID NOS: 4484
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2894
 ; LENGTH: 3495
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (88)..(840)
 ; US-09-611-526-2894

Query Match 22.3%; Score 476.8; DB 23; Length 3495;
 Best Local Similarity 87.1%; Pred. No. 7e-67;
 Matches 603; Conservative 0; Mismatches 7; Indels 82; Gaps 4;
 QY 8 ggtcaaggcgaagctcgcgggggacagcagcggggagctcctcggggagtagtacc 67
 DB 321 ggtcaaggcgaagctcgcgggggacagcagcggggagctcctcggggagtagtacc 380
 QY 68 cgggattccagagctcagcgcgtgagacgtcgcgctcctcgagcccccagcgcgc 127
 DB 381 cgggattccagagctcagcgcgtgagacgtcgcgctcctcgagcccccagcgcgc 440
 QY 128 cgctgcaacgtgcaactcctcgtcgtctgctacccgcgcagccgcgcgcgggtgtt 187
 DB 441 cgctgcaacgtgcaactcctcgtcgtctgctacccgcgcagccgcgcgcgggtgtt 500
 QY 188 ttgcccctgggcgcgtggctcgtggaaggagcctccaccggcgctccgcatgacc 247
 DB 501 ---gcccctgggcgcgtggctcgtggaaggagcctccaccggcgctccgcatgacc 556
 QY 248 cagtgaaatcaaggtgaagcagtggtgctactacacagtaataatccggaagaagcaact 307
 DB 557 cagtgaaatcaagg-aagcagggtgctactacacagtaataatccggaagaagcaact 615
 QY 308 ttgcagaatctcttgcaggaatcctgtgcaagtcccatggtcccgaggaactagag 367
 DB 616 ttgcagaatctcttgcaggaatcctgtgcaaatcccatggtcccgaggaactagag 675
 QY 368 gatgcctcctgctgtctcttaagaagattccaccgaatggtgatatgcaattgaaa 427
 DB 676 gatgcctcctgctgtctcttaagaagattccaccgaatggtgatatgcaattgaaa 735
 QY 428 gggggcacacaaatgctatgtagacaattctagacaagaagaactaaaagcactccat 487
 DB 736 gggggcacacaaatgctatgtagacaattctagacaagaagaactaaaagcactccat 795
 QY 488 ttggttctcagtagcaagatcaaaaataattatctacagtcagatgcctcaaaccaatg 547
 DB 796 ttggttctcagtagcaagatcaaaaataattatctacagtcag- 838
 QY 548 actgcttagtagggagatttttgcagcatcaacaaaataattatctacacaaacaa 607
 DB 839 ----- 838
 QY 608 cttagggagccttaccatcgttagtcaacgggtgctgtcttccctcgggagtagtacc 667
 DB 839 cttagggagccttaccatcgttagtcaacgggtgctgtcttccctcgggagtagtacc 898
 QY 668 ccaggaccaccacaaaataactttggtgctggta 699

Db 899 ccaggaccaccacaaaataactttggctgggta 930
 RESULT 8
 PCT-US01-01329-425
 ; Sequence 425, Application PC/TUS0101329
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc., et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PAI20PCT
 ; CURRENT APPLICATION NUMBER: PCT/US01/01329
 ; CURRENT FILING DATE: 2001-01-17
 ; NUMBER OF SEQ ID NOS: 3506
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 425
 ; LENGTH: 445
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (431)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (438)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; PCT-US01-01329-425

Query Match 20.1%; Score 429; DB 1; Length 445;
 Best Local Similarity 98.9%; Pred. No. 2.6e-59;
 Matches 440; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
 QY 228 ccggggcgtcccgatgatcccgatgaaatcaaggtgaagcaggtgctactactacaagta 287
 DB 2 ccggggcgtcccgatgatcccgatgaaatcaaggtgaagcaggtgctactactacaagta 60
 QY 288 ataactccggaaagaacactttgcagaatctcttgcaggaatcttgcaggaatcttgcaggttcc 347
 DB 61 ataactccggaaagaacactttgcagaatctcttgcaggaatcttgcaggttcc 120
 QY 348 catggtcccgaggaactagagatgctcctcgtgtcttcttaagaagaattccaccacaa 407
 DB 121 catggtcccgaggaactagagatgctcctcgtgtcttcttaagaagaattccaccacaa 180
 QY 408 tggatgatgccaattgaaagggggcacacaaatgctatgtagacaattcttagacaa 467
 DB 181 tggatgatgccaattgaaagggggcacacaaatgctatgtagacaattcttagacaa 240
 QY 468 gagaactaaaagcactccatttgggttcctcagtagatcaagaatcaaaaataattctcag 527
 DB 241 gagaactaaaagcactccatttgggttcctcagtagatcaagaatcaaaaataattctcag 300
 QY 528 cagatgtccctaaaacaaatgactgtctttagaggagattttgcagcatcaacaaaat 587
 DB 301 cagatgtccctaaaacaaatgactgtctttagaggagattttgcagcatcaacaaaat 360
 QY 588 taatctcattacacaaacttgaggagccttaccatcggtagtagcaggggtctgctt 647
 DB 361 taatctcattacacaaacttgaggagccttaccatcggtagtagcaggggtctgctt 420
 QY 648 tccctcgggagtagcactcttccagg 672
 DB 421 tccctcgggtagcactcttccagg 445

RESULT 9
 PCT-US01-01339-1301
 ; Sequence 1301, Application PC/TUS0101339
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc., et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006PCT


```
Matches 440; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 228 ccgcgccgctcccatgatcccgatgaaatcaaggtaaagcaggtggtactactacaagta 287
Db 2 ccgcgccgctcccatgatcccgatgaaatcaaggtaaagcaggtggtactactacaagta 60
QY 288 ataatacggagaagaagcaactttgcagaatcttctgctcaggaaatcctgttgcgaagtcc 347
Db 61 ataatacggagaagaagcaactttgcagaatcttctgctcaggaaatcctgttgcgaagtcc 120
QY 348 catgtgccaggaactagaggtgctcctgctgtcttcttaagaagaattccaacccaa 407
Db 121 catgtgccaggaactagaggtgctcctgctgtcttcttaagaagaattccaacccaa 180
QY 408 tggatgatgccaattgaaagggggcacacaatgctatgtatagacaattctagaacaa 467
Db 181 tggatgatgccaattgaaagggggcacacaatgctatgtatagacaattctagaacra 240
QY 468 gagaactaaaagcactccacttgggttcctcagatcaagatcaaaataattatctacagt 527
Db 241 gagaactaaaagcactccacttgggttcctcagatcaagatcaaaataattatctacagt 300
QY 528 cagatgtccctaaaacaaatgaactgttttagtagggagattttgcagcatcaacaaaaat 587
Db 301 cagatgtccctaaaacaaatgaactgttttagtagggagattttgcagcatcaacaaaaat 360
QY 588 taaatctattacacaacaaacttgaggagcctaccatcgtagtcaacgggtctctgt 647
Db 361 taaatctattacacaacaaacttgaggagcctaccatcgtagtcaacgggtctctgt 420
QY 648 tccctcgggatacaactcttcagg 672
Db 421 tccctcgggntcaactnttcagg 445

RESULT 12
US-09-362-510-45961
; Sequence 45961, Application US/09362510
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-759CON1
; CURRENT APPLICATION NUMBER: US/09/362,510
; EARLIER FILING DATE: 1999-07-27
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 62165
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45961
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-362-510-45961

Query Match 18.9%; Score 404; DB 17; Length 465;
Best Local Similarity 98.3%; Pred. No. 2.7e-55;
Matches 415; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 1150 ggccacatgcgctgctgtctgctcaggagaagggccgagaagaacactgctccaa 1209
Db 45 ggccacatgcgctgctgtctgctcaggagaagggccgagaagaacactgctccaa 104
QY 1210 gfgcctgcagacagatgctcctgaggaatttggaaggtcttatcacagattctcca 1269
Db 105 -nnnnnngcagagcagatgctcctggaggaatttggaaggtcttatcacagattctcca 163
QY 1270 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 1329
Db 164 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 223
QY 1330 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 1389
Db 224 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 283
QY 1390 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 1449
Db 284 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 343

Query Match 18.9%; Score 404; DB 17; Length 465;
Best Local Similarity 98.3%; Pred. No. 2.7e-55;
Matches 415; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 1150 ggccacatgcgctgctgtctgctcaggagaagggccgagaagaacactgctccaa 1209
Db 45 ggccacatgcgctgctgtctgctcaggagaagggccgagaagaacactgctccaa 104
QY 1210 gfgcctgcagacagatgctcctgaggaatttggaaggtcttatcacagattctcca 1269
Db 105 -nnnnnngcagagcagatgctcctggaggaatttggaaggtcttatcacagattctcca 163
QY 1270 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 1329
Db 164 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 223
QY 1330 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 1389
Db 224 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 283
QY 1390 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 1449
Db 284 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 343
```

```
Db 164 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 223
QY 1330 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 1389
Db 224 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 283
QY 1390 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 1449
Db 284 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 343
QY 1450 tctctgaagatggtgcaacccatgccccctttcttgaggaggtgcatgagcctgagcattg 1509
Db 344 tctctgaagatggtgcaacccatgccccctttcttgaggaggtgcatgagcctgagcattg 403
QY 1510 ttgtctggccagagagagagcttggttcccatagtcctcggagagatgtctcaggg 1569
Db 404 ttgtctggccagagagagagcttggttcccatagtcctcggagagatgtctcaggg 463
QY 1570 cg 1571
Db 464 cg 465

RESULT 13
US-09-362-510A-45961
; Sequence 45961, Application US/09362510A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-759CON1
; CURRENT APPLICATION NUMBER: US/09/362,510A
; CURRENT FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 09/221,820
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 62165
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45961
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-362-510A-45961

Query Match 18.9%; Score 404; DB 17; Length 465;
Best Local Similarity 98.3%; Pred. No. 2.7e-55;
Matches 415; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 1150 ggccacatgcgctgctgtctgctcaggagaagggccgagaagaacactgctccaa 1209
Db 45 ggccacatgcgctgctgtctgctcaggagaagggccgagaagaacactgctccaa 104
QY 1210 gfgcctgcagacagatgctcctgaggaatttggaaggtcttatcacagattctcca 1269
Db 105 -nnnnnngcagagcagatgctcctggaggaatttggaaggtcttatcacagattctcca 163
QY 1270 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 1329
Db 164 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 223
QY 1330 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 1389
Db 224 gattttgtcttagctagaatactcttagtttagaagagatgttttaggggaacatgaggt 283
QY 1390 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 1449
Db 284 ggcctgcagacaacacagcgtccctgcattccctgagggaggttttactcagagc 343
```


Search completed: July 11, 2002, 12:28:58
Job time: 15225 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 12:40:15 ; Search time 690.24 Seconds
(without alignments)
5205.529 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aattcggggtcaggcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1156321 seqs, 841861299 residues

Total number of hits satisfying chosen parameters: 2312642

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending_Patents_NA_New : *
- 1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq : *
 - 2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq : *
 - 3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq : *
 - 4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq : *
 - 5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq : *
 - 6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq : *
 - 7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2127.6	99.7	2134	1	PCT-US02-10421-2790
2	2127.6	99.7	2134	6	US-10-112-699-2790
3	2127.6	99.7	2134	6	US-10-112-699-2790
4	483.4	22.7	486	1	PCT-US02-10421-2370
5	483.4	22.7	486	6	US-10-112-699-2370
6	245	11.5	250	6	US-10-106-698-2655
7	200.2	9.4	352	5	US-09-721-544-15843
8	93.6	4.4	1684	6	US-10-155-881-3628
9	66.4	3.1	1254	6	US-10-155-881-3943
10	57	2.7	900	6	US-10-155-881-3945
11	56	2.6	530	6	US-10-027-632-221356
12	52	2.4	986	1	PCT-US02-01109-45
13	52	2.4	1772	6	US-10-102-806-196
14	51.4	2.4	381	6	US-10-138-145-301
15	49.4	2.3	763	6	US-10-155-881-3101
16	48.4	2.3	249	5	US-09-975-254-20153
17	48.4	2.3	7332	5	US-09-053-375B-1480
18	48.4	2.3	7332	5	US-09-442-384B-519
19	48.4	2.3	7693	6	US-10-160-497-11
20	48.4	2.3	55001	6	US-10-160-497-4
21	48.2	2.3	316	6	US-10-102-524-1564
22	47.8	2.2	829	6	US-10-155-881-383
23	47.6	2.2	2607	6	US-10-106-698-569
24	47.4	2.2	640	1	PCT-US02-12612-896
25	47.4	2.2	640	6	US-10-125-968-896
26	47.4	2.2	2692	5	US-09-919-002-946

ALIGNMENTS

RESULT 1

PCT-US02-10421-2790

; Sequence 2790, Application PC/TUS0210421

; GENERAL INFORMATION:

; APPLICANT: Corixa Corporation

; APPLICANT: Wang, Tongtong

; APPLICANT: Wang, Siging

; APPLICANT: Bangor, Chaitanya S.

; APPLICANT: Gaiger, Alexander

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.565PC

; CURRENT APPLICATION NUMBER: PCT/US02/10421

; CURRENT FILING DATE: 2002-03-28

; NUMBER OF SEQ ID NOS: 2959

; SEQ ID NO 2790

; LENGTH: 2134

; TYPE: DNA

; ORGANISM: Homo sapiens

PCT-US02-10421-2790

Query Match 99.7%; Score 2127.6; DB 1; Length 2134;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2130; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	1	aattcggggtcaggcgaagtcgcggggggcgacagcagcgggggagctctcggggg	60
Db	1	aattcggggtcaggcgaagtcgcggggggcgacagcagcgggggagctctcggggg	60
Qy	61	agtaccgcggatcccaagtcagcgcgtggaagagctgcgctcctcagggccgcgc	120
Db	61	agtaccgcggatcccaagtcagcgcgtggaagagctgcgctcctcagggccgcgc	120
Qy	121	agccgcgcgcctgcaacgtgcacttcctctcgtcgtcctacccgcgcagccgcgc	180
Db	121	agccgcgcgcctgcaacgtgcacttcctctcgtcgtcctacccgcgcagccgcgc	180
Qy	181	gggtgttttgcctcggggcgccctggttcctcgaaggagctcccccggggcgctcgc	240
Db	181	gggtgttttgcctcggggcgccctggttcctcgaaggagctcccccggggcgctcgc	240
Qy	241	atgacccagttgaaatacaggtaagcaggctgggtactactacaagtaataatccgaaga	300
Db	241	atgacccagttgaaatacaggtaagcaggctgggtactactacaagtaataatccgaaga	300
Qy	301	agcaactttgcagaatcttcttgcaggaaatcctgttgcaggttcccatggtccaggga	360
Db	301	agcaactttgcagaatcttcttgcaggaaatcctgttgcaggttcccatggttccaggga	360

QY 361 actagagatgacctcctgctgttctctttaagaaagattccaaaccccaatggtgatagcca 420
 Db 361 actagagatgacctcctgctgttctctttaagaaagattccaaaccccaatggtgatagcca 420
 QY 421 attgaaagggggcacacaaatgtatgtatagacaattctagacaagagaactaaagc 480
 Db 421 attgaaagggggcacacaaatgtatgtatagacaattctagacaagagaactaaagc 480
 QY 481 actccatttgggttccctcaagtatacaagatacaaaataaattatctacagtcagatccctaa 540
 Db 481 actccatttgggttccctcaagtatacaagatacaaaataaattatctacagtcagatccctaa 540
 QY 541 accaatgactgcttttagtagggagattttgcccagatcaacaaaaatlaaatctcattac 600
 Db 541 accaatgactgcttttagtagggagattttgcccagatcaacaaaaatlaaatctcattac 600
 QY 601 acaacaacttgaggagccttaccatcggtagtagcaacgggtctgctttccctcgggac 660
 Db 601 acaacaacttgaggagccttaccatcggtagtagcaacgggtctgctttccctcgggac 660
 QY 661 aactttccaggaccacaaataaactttggtggtgactgtgactgcttttccagtg 720
 Db 661 aactttccaggaccacaaataaactttggtggtgactgtgactgcttttccagtg 720
 QY 721 ggacttttgcaacaactgcaattgtaataattgttgcaacaacttgcatcatgatattga 780
 Db 721 ggacttttgcaacaactgcaattgtaataattgttgcaacaacttgcatcatgatattga 780
 QY 781 acggtttaagccattaaagcattgcttggtagaataccagagccttccagccaaat 840
 Db 781 acggtttaagccattaaagcattgcttggtagaataccagagccttccagccaaat 840
 QY 841 tgggaagggccaatggcgaattgcaagcccgacacaaagggtgcaactgcaggag 900
 Db 841 tgggaagggccaatggcgaattgcaagcccgacacaaagggtgcaactgcaggag 900
 QY 901 gtcaggctgctgaagaatttactgcagtgctgtatgagcccaaatattgtgttctctat 960
 Db 901 gtcaggctgctgaagaatttactgcagtgctgtatgagcccaaatattgtgttctctat 960
 QY 961 ttgcaaatgcatggttgcacaaatattgaagaagcccgagaaagacactaatag 1020
 Db 961 ttgcaaatgcatggttgcacaaatattgaagaagcccgagaaagacactaatag 1020
 QY 1021 catgccaaactacatgcagactgaggttggaggcagcattacctgccacacgaa 1080
 Db 1021 catgccaaactacatgcagactgaggttggaggcagcattacctgccacacgaa 1080
 QY 1081 attttcaggacttccaagattcagtcagataggcgcccttccatcatgcatctctggga 1140
 Db 1081 attttcaggacttccaagattcagtcagataggcgcccttccatcatgcatctctggga 1140
 QY 1141 ggtggtgagggcacatgcgcttgcctgctgtgtctcaggggagagagcccgagaaagaa 1200
 Db 1141 ggtggtgagggcacatgcgcttgcctgctgtgtctcaggggagagagcccgagaaagaa 1200
 QY 1201 ctgctccaagtgcctggcagagatgctctggaggaatttggaaagtgcattacaa 1260
 Db 1201 ctgctccaagtgcctggcagagatgctctggaggaatttggaaagtgcattacaa 1260
 QY 1261 gatttcccaactgagtttaaatcttaaggagattgaaatggagtagtagtaaaagtga 1320
 Db 1261 gatttcccaactgagtttaaatcttaaggagattgaaatggagtagtagtaaaagtga 1320
 QY 1321 atgcatgtatttcttagttagtaaaatctctagtttagtaaaagagattttaggggaa 1380
 Db 1321 atgcatgtatttcttagttagtaaaatctctagtttagtaaaagagattttaggggaa 1380
 QY 1381 catgaggtgctctgcagcaacacacaggtctccctgcacccctcctggcccgaggagttt 1440
 Db 1381 catgaggtgctctgcagcaacacacaggtctccctgcacccctcctggcccgaggagttt 1440

QY 1441 actcagagctctctgaagatgtggcaaccccatgcccccttttctgaggaggtgcatggcc 1500
 Db 1441 actcagagctctctgaagatgtggcaaccccatgcccccttttctgaggaggtgcatggcc 1500
 QY 1501 tgagcatttgttctggtgcccagagagagagccttgggttcccatagtcctcggagagtg 1560
 Db 1501 tgagcatttgttctggtgcccagagagagagccttgggttcccatagtcctcggagagtg 1560
 QY 1561 tctgcaggcggcgagggcagagcagcctcgcgagagctcactctcgtcgaactctc 1620
 Db 1561 tctgcaggcggcgagggcagagcagcctcgcgagagctcactctcgtcgaactctc 1620
 QY 1621 ctctcagaataattgtctctgaggctgctcgtcatgaaacccctaatggtttcttgtt 1680
 Db 1621 ctctcagaataattgtctctgaggctgctcgtcatgaaacccctaatggtttcttgtt 1680
 QY 1681 gtttttcaaatattttagaataaagtctccggatgggtggttctgataccacttataat 1740
 Db 1681 gtttttcaaatattttagaataaagtctccggatgggtggttctgataccacttataat 1740
 QY 1741 ctctagagaactactgaacacccctaaagatttctgtagcgtagatatttcccagagaca 1800
 Db 1741 ctctagagaactactgaacacccctaaagatttctgtagcgtagatatttcccagagaca 1800
 QY 1801 cgcgaactgtcagctcttcttaaggccccgggagacgagcagcaatggggcctcgcaggc 1860
 Db 1801 cgcgaactgtcagctcttcttaaggccccgggagacgagcagcaatggggcctcgcaggc 1860
 QY 1861 caggcttgaccagatgctgtgagttagtaggagacttaaaattatccagtttcttctgt 1920
 Db 1861 caggcttgaccagatgctgtgagttagtaggagacttaaaattatccagtttcttctgt 1920
 QY 1921 ttctactgaattgtgaaagctctattccaaatcttccatctccataattattgtgt 1980
 Db 1921 ttctactgaattgtgaaagctctattccaaatcttccatctccataattattgtgt 1980
 QY 1981 aatattattattgttgtgaaacatggttccataactagctgttggaacccagcaggtta 2040
 Db 1981 aatattattattgttgtgaaacatggttccataactagctgttggaacccagcaggtta 2040
 QY 2041 aaatgaattcttaagtgtgacgttttgggttctgttgaagcaaatgaataaaattt 2100
 Db 2041 aaatgaattcttaagtgtgacgttttgggttctgttgaagcaaatgaataaaattt 2100
 QY 2101 ccaatgtcgaacaaaaaaaccccaaaaaaa 2134
 Db 2101 ccaatgtcgaacaaaaaaaccccaaaaaaa 2134

RESULT 2

US-10-112-699-2790
 ; Sequence 2790, Application US/10112699

GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong
 ; APPLICANT: Wang, Siqing
 ; APPLICANT: Bangur, Chaitanya S.
 ; APPLICANT: Gaiger, Alexander
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF CANCER
 ; FILE REFERENCE: 210121.565
 ; CURRENT APPLICATION NUMBER: US/10/112,699
 ; CURRENT FILING DATE: 2002-03-28
 ; NUMBER OF SEQ ID NOS: 2959
 ; SEQ ID NO 2790
 ; LENGTH: 2134
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-112-699-2790

Query Match 99.7%; Score 2127.6; DB 6; Length 2134;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2130; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

RESULT 3

US-10-172-118-1104
: Sequence 1104, Application US/10172118

: GENERAL INFORMATION:

: APPLICANT: Dai, Hongyue
: APPLICANT: He, Yudong
: APPLICANT: Linsley, Peter
: APPLICANT: Mao, Mao
: APPLICANT: Roberts, Chris
: APPLICANT: Van 't Veer, Laura
: APPLICANT: Van de Vijver, Marc

: APPLICANT: Bernards, Rene

: TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients

: FILE REFERENCE: 9301-175-999

: CURRENT APPLICATION NUMBER: US/10/172.118

: CURRENT FILING DATE: 2002-06-14

: PRIOR APPLICATION NUMBER: 60/380,770

: PRIOR FILING DATE: 2002-05-14

: NUMBER OF SEQ ID NOS: 2699

: SEQ ID NO 1104

: LENGTH: 2134

: TYPE: DNA

: ORGANISM: Homo sapiens

: PUBLICATION INFORMATION:

: DATABASE ACCESSION NUMBER: NM_004923

: DATABASE ENTRY DATE: 2001-06-18

US-10-172-118-1104

Query Match 99.7%; Score 2127.6; DB 6; Length 2134;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2130; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	1	aattcggggtcaaggcagctcgcgggggacagcgacgggagggagctctcggg	60
Dd	1	aattcggggtcaaggcagctcgcgggggacagcgacgggagggagctctcggg	60
Qy	61	agtaccgcggatccagagctcgcgcgtggaggagctgcgtctcgcagcccg	120
Dd	61	agtaccgcggatccagagctcgcgcgtggaggagctgcgtctcgcagcccg	120
Qy	121	agccgcgcgcgcgaacgtgcactctctcgtctcgtcgtcgtcgtcgtcgtc	180
Dd	121	agccgcgcgcgcgaacgtgcactctctcgtctcgtcgtcgtcgtcgtcgtc	180
Qy	181	gggtgttttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	240
Dd	181	gggtgttttgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	240
Qy	241	atgattccaggtgaaatcaagtgtaagcagggtggtactactacaaagtaata	300
Dd	241	atgattccaggtgaaatcaagtgtaagcagggtggtactactacaaagtaata	300
Qy	301	agcaactttgcagaatcttctgtcaggaatcctgttgcgaagtcccatggtcc	360
Dd	301	agcaactttgcagaatcttctgtcaggaatcctgttgcgaagtcccatggtcc	360
Qy	361	actagagatccctcgtctgtcttctaagaagattccaaaccatggtgatgcca	420
Dd	361	actagagatccctcgtctgtcttctaagaagattccaaaccatggtgatgcca	420
Qy	421	attgaagggggacacaaatgctatgtatagacaaatcttagaacaagaaact	480
Dd	421	attgaagggggacacaaatgctatgtatagacaaatcttagaacaagaaact	480
Qy	481	actccattgttctcagtaacaaatgctatgtatagacaaatcttagaacaag	540
Dd	481	actccattgttctcagtaacaaatgctatgtatagacaaatcttagaacaag	540
Qy	541	accaataactcttagtagggagattttgcagcagcatcaacaaatcaatcttac	600
Dd	541	accaataactcttagtagggagattttgcagcagcatcaacaaatcaatcttac	600

Qy	601	acaacaacttgaggagccttaccatcggtagtcacacgggtctgcttccctcgggac	660
Dd	601	acaacaacttgaggagccttaccatcggtagtcacacgggtctgcttccctcgggac	660
Qy	661	aactctccaggaccacacaaataaactttggctgggtactgtgactgcttggcagtg	720
Dd	661	aactctccaggaccacacaaataaactttggctgggtactgtgactgcttggcagtg	720
Qy	721	ggacttttgcacaactgcgaattgttaataattgttgcacaacttgcatacatatga	780
Dd	721	ggacttttgcacaactgcgaattgttaataattgttgcacaacttgcatacatatga	780
Qy	781	acggtttaagccattaaagcattgcttggtagaaatccagaagctttccagcacaat	840
Dd	781	acggtttaagccattaaagcattgcttggtagaaatccagaagctttccagcacaat	840
Qy	841	tgggaaggcccaatttggccaattgtcaagcccccacacacaaagggtgcactgcag	900
Dd	841	tgggaaggcccaatttggccaattgtcaagcccccacacacaaagggtgcactgcag	900
Qy	901	gtcaggtcgtcgaagaatttactgcaggtgtatgagggcccaaatattgttctctat	960
Dd	901	gtcaggtcgtcgaagaatttactgcaggtgtatgagggcccaaatattgttctctat	960
Qy	961	tgcacaatgcattggttgcacaatattgaagaagccagacgacgaagacacataag	1020
Dd	961	tgcacaatgcattggttgcacaatattgaagaagccagacgacgaagacacataag	1020
Qy	1021	catgccaaatacatgcagactggaggttgggaaggcagccattaccctgcacacaa	1080
Dd	1021	catgccaaatacatgcagactggaggttgggaaggcagccattaccctgcacacaa	1080
Qy	1081	atttcaggacttccagattcagtcagatagcgccttccctcctcctcctcctgga	1140
Dd	1081	atttcaggacttccagattcagtcagatagcgccttccctcctcctcctcctgga	1140
Qy	1141	gggtggtggagccacatgcgcctgcctgcttgcagggagagagagccgagaaaga	1200
Dd	1141	gggtggtggagccacatgcgcctgcctgcttgcagggagagagagccgagaaaga	1200
Qy	1201	ctgctccaaagtgcctggcagagcagatgacctggaggaatttgggaaggtgcttaca	1260
Dd	1201	ctgctccaaagtgcctggcagagcagatgacctggaggaatttgggaaggtgcttaca	1260
Qy	1261	gattctccacactgagtttaaatcgaaggattgaataatgagtagagataaaagtga	1320
Dd	1261	gattctccacactgagtttaaatcgaaggattgaataatgagtagagataaaagtga	1320
Qy	1321	atgcattgtgattttgtcttagtctagaaatctctagtttagaaagagatgttaggg	1380
Dd	1321	atgcattgtgattttgtcttagtctagaaatctctagtttagaaagagatgttaggg	1380
Qy	1381	catgaggtcgtcctgcagcaaacacccagcctccctgcacccctggcccgaggagtt	1440
Dd	1381	catgaggtcgtcctgcagcaaacacccagcctccctgcacccctggcccgaggagtt	1440
Qy	1441	actcagagctctcgaagatgtggcaaccatgcccttttctgagagagtgcatggcc	1500
Dd	1441	actcagagctctcgaagatgtggcaaccatgcccttttctgagagagtgcatggcc	1500
Qy	1501	tgagcattgttctgctggccagagagagcttgggttcccaatagctcctgggagagtg	1560
Dd	1501	tgagcattgttctgctggccagagagagagcttgggttcccaatagctcctgggagagtg	1560
Qy	1561	tcctcagggcgccggagggcagagcagccctgggagagctcactgtgtcagactctc	1620
Dd	1561	tcctcagggcgccggagggcagagcagccctgggagagctcactgtgtcagactctc	1620
Qy	1621	ctctcagagaatgtgtcctcggaggtcctgcataaaacccctaatggtttcttcttt	1680
Dd	1621	ctctcagagaatgtgtcctcggaggtcctgcataaaacccctaatggtttcttcttt	1680
Qy	1681	gttttcaaatatttagaataaagtctccggaggttctcggaggtgtgtgtgatacaactaaat	1740

Db	1681	gtttttccaattattctagaaa	caagttctccggatg	gggtgtgttgataccaacttaaat	1740
Qy	1741	ctctagagaactactgacaa	cacctaaagattttctgt	tagcgtagatatctccagagaca	1800
Db	1741	ctctagagaactactgacaa	cacctaaagattttctgt	tagcgtagatatctccagagaca	1800
Qy	1801	cgcgaactgtcagttctt	ctaaggccccggagacg	caggcaatggggctcgcaggc	1860
Db	1801	cgcgaactgtcagttctt	ctaaggccccgggagacg	caggcaatggggctcgcaggc	1860
Qy	1861	caggcttgaccagacgat	gtcttgagttgaggaact	taaaattatccagttttctctgt	1920
Db	1861	caggcttgaccagacgat	gtcttgagttgaggaact	taaaattatccagttttctctgt	1920
Qy	1921	tctcacttgaaatgtgg	aaaaagctctattccaat	taacttccataattattgtgt	1980
Db	1921	tctcacttgaaatgtgg	aaaaagctctattccaat	taacttccataattattgtgt	1980
Qy	1981	aatattattatgtttgt	taaaacatggttcacata	aactagcttgttgaaaccagcaggt	2040
Db	1981	aatattattatgtttgt	taaaacatggttcacata	aactagcttgttgaaaccagcaggt	2040
Qy	2041	aaatgaattctttaa	gtgacgcgtttttgtct	gtttgtgttaagcaagaatgaataaaattt	2100
Db	2041	aaatgaattctttaa	gtgacgcgtttttgtct	gtttgtgtgttaagcaagaatgaataaaattt	2100
Qy	2101	ccaatgtcga	aaaaaaaaaaaaaaaaaaaa	2134	
Db	2101	ccaatgtcga	aaaaaaaaaaaaaaaaaaaa	2134	

```

RESULT      4
PCT-US02-10421-2370
; Sequence 2370, Application PC/TUS0210421
; GENERAL INFORMATION:
; APPLICANT: Corixa Corporation
; APPLICANT: Wang, Tongtong
; APPLICANT: Wang, Siqing
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Galger, Alexander
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF CANCER
; FILE REFERENCE: 210121.565PC
; CURRENT APPLICATION NUMBER: PCT/US02/10421
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 2959
; SEQ ID NO 2370
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 484
; OTHER INFORMATION: n = A,T,C or G
PCT-US02-10421-2370

```

[illegible]

Qy	482	ctccatttggttcctcagtagtcaagatcaaaaataattatctacagtcagatgccccataa	541
Db	181	ctccatttggttcctcagtagtcaagatcaaaaataattatctacagtcagatgccccataa	240
Qy	542	ccaatgactgcttttagtagggagatttttgcagagatcaacaaaaataaatctcattaca	601
Db	241	ccaatgactgcttttagtagggagatttttgcagagatcaacaaaaataaatctcattaca	300
Qy	602	caacaacttgaggaggagccttaccatcggtagtcaacgggctctgctttccctcoggagatca	661
Db	301	caacaacttgaggaggagccttaccatcggtagtcaacgggctctgctttccctcoggagatca	360
Qy	662	actcttcaggagaccacaaaaataactttggctgggtactgtgactgctttgcagatggg	721
Db	361	actcttcaggagaccacaaaaataactttggctgggtactgtgactgctttgcagatggg	420
Qy	722	gacttttgcaacaactgcaattgtaataattgttgcaacaacttgcatcatgatattgaa	781
Db	421	gacttttgcaacaactgcaattgtaataattgttgcaacaacttgcatcatgatattgaa	480
Qy	782	cggttt	787
Db	481	cgqntt	486

```

RESULT      5
US-10-112-699-2370
; Sequence 2370, Application US/10112699
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Wang, Siqing
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Gaiger, Alexander
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF CANCER
; FILE REFERENCE: 210121.565
; CURRENT APPLICATION NUMBER: US/10/112,699
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 2959
; SEQ ID NO 2370
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 484
; OTHER INFORMATION: n = A,T,C or G
US-10-112-699-2370

```

	Query Match	22.7%;	Score 483.4;	DB 6;	Length 486;
	Best Local Similarity	99.6%;	Prod. No. 9e-83;		
	Matches 484;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	302	gcaacttgcagaatcttctgctcagaaatccctgtgtcgaagtcccatgtccaggaa	361		
Db	1	gcaacttgcagaatcttctgctcagaaatccctgtgtcgaagtcccatgtccaggaa	60		
Qy	362	ctagagatgctcctgctgtctcttaagaaaagatcccaaccctgaatgtgatgcaa	421		
Db	61	ctagagatgctcctgctgtctcttaagaaaagatcccaaccctgaatgtgatgcaa	120		
Qy	422	ttagaagggggcacacaaatgctgtatagacaaattctagacaaagagaactaaagca	481		
Db	121	ttagaagggggcacacaaatgctgtatagacaaattctagacaaagagaactaaagca	180		
Qy	482	ctccatttggcttcctcagatccaagaataattatctacagtcagatgtccctaaa	541		
Db	181	ctccatttggcttcctcagatccaagaataattatctacagtcagatgtccctaaa	240		
Qy	542	ccaatgactgctttagtgaggagattttgccagcatcaacaaataataatctcattaca	601		

Db 241 ccaatgactgttttagtgaggagatttttgcagcatcaacaaataaataatctcattaca 300
QY 602 caacaactgaggagccttaccatcggtagtcacacgggtctgcttccctccgagatca 661
|||||
Db 301 caacaactgaggagccttaccatcggtagtcacacgggtctgcttccctccgagatca 360
QY 662 actcttccaggaccaccacaaataaacttggctgggtactgtgactgtttgcccagtggg 721
|||||
Db 361 actcttccaggaccaccacaaataaacttggctgggtactgtgactgtttgcccagtggg 420
QY 722 gactttgcaacaaactgcaattgtaataattgttgcacaaacttgcacatgatattgaa 781
|||||
Db 421 gactttgcaacaaactgcaattgtaataattgttgcacaaacttgcacatgatattgaa 480
QY 782 cggttt 787
||| ||
Db 481 cggntt 486

RESULT 6
US-10-106-698-2655
; Sequence 2655, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO 2655
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (199)..(199)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (204)..(204)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (214)..(214)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (232)..(232)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (243)..(243)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-2655

Query Match 11.5%; Score 245; DB 6; Length 250;
Best Local Similarity 98.0%; Pred. No. 1.5e-37;
Matches 245; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1144 ggtgagagccacatgcctgcctgctctcaggagagagccgagagaaagaaactg 1203
|||||
Db 1 ggtgagagccacatgcctgcctgctctcaggagagagccgagagaaagaaactg 60
QY 1204 ctccaagtgcctggcagagcagatgacccctggaggaatttgggaaggtgcttatcacagat 1263
|||||
Db 61 ctccaagtgcctggcagagcagatgacccctggaggaatttgggaaggtgcttatcacagat 120
QY 1264 tctccacactgagtttaaatctaaaggagattgaaatggagtagagataaagtgtaag 1323
|||||

Db 121 tctccacactgagtttaaaatctaaaggattgaaaatggagtagagtataaaagtgtgaatg 180
QY 1324 catgttgaatttgccttagtctagaataatctctagtttagaagaagatgttttaggggaacat 1383
|||||
Db 181 catgttgaatttgccttagtctagaataatctctagtttagaagaagatgttttaggggaacat 240
QY 1384 gaggctggct 1393
||| |||||
Db 241 gaggctggct 250

RESULT 7
US-09-721-544-15843
; Sequence 15843, Application US/09721544
; GENERAL INFORMATION:
; APPLICANT: Arterburn, Matthew
; APPLICANT: Asghari, Vida
; APPLICANT: Damavandi, Simin
; APPLICANT: Dickson, Mark
; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Garcia, Veronica
; APPLICANT: Giedt, Gretchen
; APPLICANT: Hunter, Kelly
; APPLICANT: Jensen, Aaron
; APPLICANT: Jones, Lee
; APPLICANT: Kita, David
; APPLICANT: Labat, Ivan
; APPLICANT: Laroya, Mimi
; APPLICANT: Lomelli, Michelle
; APPLICANT: Nguyen, Phuong
; APPLICANT: Nogra, Margie
; APPLICANT: Palencia, Servando
; APPLICANT: Raisi, Fariba
; APPLICANT: Smith, Benjamin
; APPLICANT: Tkach, Joe
; APPLICANT: Tran, Lien
; APPLICANT: Verna, Ron
; APPLICANT: Yang, Fei
; APPLICANT: Yim, Kenneth
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; TITLE OF INVENTION: From a CDNA Library of Fetal Liver-Spleen
; FILE REFERENCE: 728CIP
; CURRENT APPLICATION NUMBER: US/09/721,544
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 09/515,128
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 09/034,341
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 24489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15843
; LENGTH: 352
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(352)
; OTHER INFORMATION: n = A,T,C or G
US-09-721-544-15843

Query Match 9.4%; Score 200.2; DB 5; Length 352;
Best Local Similarity 75.1%; Pred. No. 5.2e-29;
Matches 250; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 697 gactgtgactgtttgcccagtgaggacttttgcacaaactgcaattgtaataattgttg 756
|||||
Db 4 gattgtgattgtttgcaaaatgtgaaattttgcaaaactgcaattgtactaatgttta 63
QY 757 caacaactgcatcatgatattgaaagggttttaaagccattaaagccatgtcttggtagaaa 816

```

Db 64 caacaatttggaacatgaaataagaaagcaaaagcaataaaggcatgcttgcagaaa 123
Qy 817 tccagaagcttcacagcccaaaaattgggaagggccaattgggcaatgtcagaagcccaagca 876
Db 124 tccagaagcttcacagcccaaaaattgggaagggccaattgggcaatgtcagaagcccaagca 183
Qy 877 caacaaggggtgcaactgcagggaggtcagcgtcgtcgaagaattactcgcgagtgctatga 936
Db 184 tagcaaaagggtgttaattgcaaacatcaggaatgtctttaaactactgtgaatgctatga 243
Qy 937 ggcacaaattatgtgtttcttatttgcacaaatgcattgttgcacaaattatgaagaag 996
Db 244 ggcacaaattatgtgtttctcctcaatgcacaaatgtattgctgtaagaatttttgaagaag 303
Qy 997 cccagaagaagaacactaatgagcatgccaaa 1029
Db 304 cccgaaagaagaacattgatgctttggcaga 336

```

RESULT 8

US-10-155-881-3628
; Sequence 3628, Application US/10155881

; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3628
; LENGTH: 1684
; TYPE: DNA
; ORGANISM: Zea mays subsp. mexicana
US-10-155-881-3628

```

Query Match 4.4%; Score 93.6; DB 6; Length 1684;
Best Local Similarity 57.5%; Pred. No. 1.1e-08;
Matches 208; Conservative 0; Mismatches 124; Indels 30; Gaps 1;

Qy 683 ataaacttggctgggtactgtgactgtgttgcagtgagggaactttgcacaaactgcaat 742
Db 241 aatgtttgaagtgtactgtgaatgtgttgcagtgagggaactttgcagtgagggaactttgcacaaactgcaat 300
Qy 743 tgaataattgtgcaaacactgtgactgtatgatattgaacgttttaagccattgaagca 802
Db 301 tgcacaaactgttttaataactgtgataatgaggtgcagggcgtgaggtgtgtgaagct 360
Qy 803 tgtcttggtagaataccagaagcttccagcccaaaaattgggaagggccaattgggcaat 862
Db 361 atattggatcgaataccagagcttccagcctaagattgggagcagccacttatgcat 420
Qy 863 gtcaagccccc-----agcacacaaaggggtgcac 892
Db 421 aggaataatagggtgcctagtgtatcctctttagtagggaagcacaataaagggtgccat 480
Qy 893 tgcagaggtcaggtcgtcgaagaattactgcaggtactgtgaggtcagcccaattatgtgt 952
Db 481 tgcagaaggtcaggtcgtcgaagaattactgcaggtactgtgaggtcagcccaattatgtgt 540
Qy 953 tcttctattgcaaatgctgtgttgcacaaattatgaagaagcccaagcccaagcaaga 1012
Db 541 tctgaaactgcaggtcgtcgaagaattatgaagaagcccaaggtgaagagagaagaat 600
Qy 1013 ct 1014
Db 601 ct 602

```

RESULT 9

US-10-155-881-3943
; Sequence 3943, Application US/10155881

; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3943
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Zea mays
US-10-155-881-3943

```

Query Match 3.1%; Score 66.4; DB 6; Length 1254;
Best Local Similarity 65.5%; Pred. No. 0.0016;
Matches 97; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 863 gtcaagcccccagcacacaaaggggtgcaactgcagaggtcaggtgctcgaagaattac 922
Db 9 ggcggggcaagcataaacccaggtgtcactgcaagaagtcaggtgtgtctaaagaataac 68
Qy 923 tgcagagtctatgagggcccaaaattatgttcttcttcaaatgcattggttgcaaa 982
Db 69 tgcagagtcttctcaagcgcaatactccttgcacaaagattgtaaatgtatgattgcaaa 128
Qy 983 aattatgaagaagcccgacgaaga 1010
Db 129 aactatgaaggagtgaggagctacgga 156

```

RESULT 10

US-10-155-881-3945
; Sequence 3945, Application US/10155881

; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3945
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(900)
; OTHER INFORMATION: unsure at all n locations
US-10-155-881-3945

```

Query Match 2.7%; Score 57; DB 6; Length 900;
Best Local Similarity 58.6%; Pred. No. 0.091;
Matches 99; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 682 aataacttggctgggtactgtgactgttgcaggtggggacttttgcacaaactgcaa 741
Db 601 ct 602

```

Db 614 aaaaattgtgaagatgactgtgaatgtctttcaagaactcaactgtgactgtggtgcaa 673
 Qy 742 ttgtaataattgtgcaacaaactgtcatcatgatattgaacggttttaaagcaactaaagc 801
 Db 674 cftttcaaatgtggaacattgttggaatgagaatgccagaacgaagaacttgagc 733
 Qy 802 atgtcttgtagaataccagaagcttccagccaaataattgggaagggc 850
 Db 734 aattgcgcagtaataccgtctgctgtttccagctaaataattggaaatgac 782

RESULT 11

US-10-027-632-221356
 ; Sequence 221356, Application US/10027632
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; Polymorphisms in the Human Genome
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; PRIOR FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 221356
 ; LENGTH: 530
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-10-027-632-221356

Query Match 2.6%; Score 56; DB 6; Length 530;
 Best Local Similarity 77.3%; Pred. No. 0.13;
 Matches 68; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 942 aaatatgtgtcttcttattgcaaatgcatgtgttgcaaaaattatgaagaagccag 1001
 Db 3 aaataatgttctcctaataatgcaaatgtattgctgtaagaattttgaagaagcccg 62
 Qy 1002 aacgaagaagactaatgagatgccaaa 1029
 Db 63 aaaggaagacattgatcatttgccaga 90

RESULT 12

PCR-US02-01109-45
 ; Sequence 45, Application PC/TUS0201109
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc.
 ; TITLE OF INVENTION: 50 Human Secreted Proteins
 ; FILE REFERENCE: P2016PCT2
 ; CURRENT APPLICATION NUMBER: PC/TUS02/01109
 ; CURRENT FILING DATE: 2002-01-17
 ; PRIOR APPLICATION NUMBER: US 60/262,066
 ; PRIOR FILING DATE: 2001-01-18
 ; NUMBER OF SEQ ID NOS: 206
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 45
 ; LENGTH: 986
 ; TYPE: DNA

; ORGANISM: Homo sapiens
 PCT-US02-01109-45

Query Match 2.4%; Score 52; DB 1; Length 986;
 Best Local Similarity 52.8%; Pred. No. 0.82;
 Matches 112; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 1923 ctacttgaaattgtggaagagctctattatccaaataattccataattattgtgttaa 1982
 Db 702 ctgacagatgtggaaggaatcagatcttctggtttactctggtgttaa 761
 Qy 1983 tattattattgtttgtaaaacatggttcacataactagctgtggaaccacagcaggtaaa 2042
 Db 762 aacagaatgatcacagaaaatgtttccttctgtaactggtgtaacatagaactggg 821
 Qy 2043 atgaattcttaagtgcagcgttttctgttctgtttaaagcaagaatgaataaaatttcc 2102
 Db 822 tattatagatcactttcacttttgggaatgtttgtattgaaacttaataaaactttaa 881
 Qy 2103 aatgtcgaaaaaataaaaaaataaaaaa 2134
 Db 882 catggcaaaaaaataaaaaaataaaaaa 913

RESULT 13

US-10-102-806-196
 ; Sequence 196, Application US/10102806
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA103PIC1
 ; CURRENT APPLICATION NUMBER: US/10/102,806
 ; CURRENT FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: 09/925,298
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05881
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 846
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 196
 ; LENGTH: 1772
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (2)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc_feature
 ; LOCATION: (1749)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc_feature
 ; LOCATION: (1769)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; US-10-102-806-196

Query Match 2.4%; Score 52; DB 6; Length 1772;
 Best Local Similarity 52.8%; Pred. No. 0.89;
 Matches 112; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 1923 ctacttgaaattgtggaagagctctattatccaaataactctcactatcattattgtgttaa 1982
 Db 1473 ctgacagatgtggaaggaatcagatcttctggtttactctggtgttaa 1532
 Qy 1983 tattattattgtttgtaaaacatggttcacataactagctgtggaaccacagcaggtaaa 2042
 Db 1533 aacagaatgatcacagaaaatgtttccttctgtaactggttagtgacatagaactggg 1592
 Qy 2043 atgaattcttaagtgcagcgttttctgttctgttgaagcaagaatgaataaaatttcc 2102

Db 1593 tattatagatcacttttcacttttttggaatgttttgattgaaacttaataaaacttttaa 1652
QY 2103 aatgtcgaaaaaataaaaaaataaaaaa 2134
Db 1653 catggcaaaaaaataaaaaaataaaaaa 1684

RESULT 14

US-10-138-145-301
; Sequence 301, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Gibson, John Bryan
; APPLICANT: Norris, Michael Geoffrey
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use.
; FILE REFERENCE: 11000.1058U
; CURRENT APPLICATION NUMBER: US/10/138,145
; CURRENT FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 1652
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 301
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(381)
; OTHER INFORMATION: n = A,T,C or G
US-10-138-145-301

Query Match 2.4%; Score 51.4; DB 6; Length 381;
Best Local Similarity 65.0%; Pred. NO. 0.94;
Matches 76; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 868 gccccagcacacaagaagggtgcaactgcaggaggtcaggtgcctgcctgaagaattactgcga 927
Db 253 ggctgctcataagagagaggtgcaattgcaagaagtcagctgtctcaagaataattgcga 312
QY 928 gtgctatgagggcccaaatgtgtctcttattgcaaatgcatgtgtgcaaaaa 984
Db 313 atgctatcaggagggtgtgtgctggtcctcacaaactgcagatgcgaaacttgcaaaa 369

RESULT 15

US-10-155-881-3101
; Sequence 3101, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3101
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Zea mays
US-10-155-881-3101

Query Match 2.3%; Score 49.4; DB 6; Length 763;
Best Local Similarity 66.4%; Pred. NO. 2.5;
Matches 71; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 878 aacaaagggtgcaactgcaggaggtcaggtgcctgaaataattactgcagtgctatgag 937
Db 599 aagagagggttgcactgcaggaagtcaggctgctaaagaatactctgtgatgctaccag 658
QY 938 gcccaaatattgtttcttctatttgcataatgcattggttgcaaaaa 984
Db 659 gatggactggatgttctctgttttgcgcgtgtgacgattgccagaa 705

Search completed: July 11, 2002, 12:40:34
Job time: 12441 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:13:36 ; Search time 53.06 seconds
(without alignments)
617.542 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVTCLKGGQMLCIDNCGA.....GRLCSILHIEFKSKGLKTE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1599	100.0	295	21	AA198463 Mouse testis speci
2	1251	78.2	299	21	AA198464 Human testis speci
3	435	27.2	950	22	AB862035 Drosophila melanog
4	396	24.8	438	20	AAW83392 Caenorhabditis ele
5	362	22.6	280	22	ABG17958 Novel human diagn
6	322	20.1	147	22	AAW95330 Human reproductive
7	248	15.5	53	22	AA197615 Human colon cancer
8	186	11.6	243	22	AB868888 Drosophila melanog
9	165	10.3	251	22	AAW93348 Human polypeptide,
10	110.5	6.9	732	18	AAW26642 Human RECK cancer-
11	99	6.2	2931	22	AB868229 Drosophila melanog

12	97	6.1	602	21	AA195660 Human Zntr2 protei
13	97	6.1	1987	22	AB861083 Drosophila melanog
14	96.5	6.0	359	21	AA190497 Arabidopsis thalia
15	96.5	6.0	443	21	AA190496 Arabidopsis thalia
16	96.5	6.0	472	21	AA190495 Arabidopsis thalia
17	96.5	6.0	1788	22	AB862995 Drosophila melanog
18	96.5	6.0	4547	22	AB859051 Drosophila melanog
19	96	6.0	773	22	AB859751 Drosophila melanog
20	95.5	6.0	1035	22	AB866062 Drosophila melanog
21	94	5.9	1792	21	AA198443 Human laminin 8 po
22	94	5.9	1800	21	AA198445 Human laminin 8 po
23	94	5.9	1816	21	AA198442 Human laminin 8 po
24	94	5.9	1824	21	AA198444 Human laminin 8 po
25	93.5	5.8	652	19	AA198479 Amino acid sequenc
26	93.5	5.8	652	21	AA198475 Human cell surface
27	93.5	5.8	1981	22	AB861657 Drosophila melanog
28	92.5	5.8	921	21	AA198297 Corn polycomb prot
29	92	5.8	3680	22	AB870878 Drosophila melanog
30	91.5	5.7	637	21	AA198346 Mouse cell surface
31	91.5	5.7	644	21	AA198346 Haematopoietic ste
32	91.5	5.7	644	21	AA198346 Haematopoietic ste
33	91	5.7	18	21	AA198465 Mouse testis speci
34	89.5	5.6	347	22	AA198443 Human gene 35 enco
35	89.5	5.6	361	22	AA198443 Human polypeptide
36	89.5	5.6	439	22	AA198443 Human protein sequ
37	89	5.6	1725	21	AA198443 Mouse laminin 2 ma
38	89	5.6	1725	21	AA198443 Mouse laminin 8 po
39	89	5.6	1764	10	AA198443 Primary amino acid
40	89	5.6	1776	19	AA198443 Mouse laminin Bl c
41	89	5.6	1786	21	AA198443 Mouse laminin 2 be
42	89	5.6	1786	21	AA198443 Mouse laminin 8 po
43	88.5	5.5	3843	22	AB871529 Drosophila melanog
44	88	5.5	250	21	AA198233 Metallothionein ve
45	88	5.5	466	11	AA198443 Human laminin Bl c

ALIGNMENTS

RESULT 1

AA198463
ID AA198463 standard; Protein: 295 AA.

AC AA198463;

DT 25-APR-2000 (first entry)

DE Mouse testis specific factor tesmin SEQ ID NO:4.

DE Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.

OS Mus musculus.

PN WO200004147-A1.

PD 27-JAN-2000.

PF 16-JUL-1999; 99WO-JP03859.

PR 17-JUL-1998; 98JP-0219856.

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

DR WPI; 2000-147785/13.

DR N-PSDB; AA288155, AA288157.

PT New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility -

XX Claim 1; Page 47-49; 63pp; Japanese.
 XX The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 295 AA;

Query Match 100.0%; Score 1599; DB 21; Length 295;
 Best Local Similarity 100.0%; Pred. No. 2.5e-144; Indels 0; Gaps 0;
 Matches 295; Conservative 0; Mismatches 0;

QY 1 MVTCLKGGAGMLCIDNCCGARELKALHLLPOYDDQSFPQSELPKPMTTLVGRLLPVPK 60
 DB 1 mvtclkggagmlcidncgarelkahlhlpoyddqsfpqselpkpmttlvgrllvpak 60
 QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTITLGGYDCFSGDFCNCSNLRHEL 120
 DB 61 lnlitqvdngalpsavngaappspalqgpptitlsgydcfsgdfcnscnnlrhel 120
 QY 121 ERFAKACLDNRNPEAFQPMKGRGLGAALKRHSKGCNCRSGCLKNYCEYEAQIMCSS 180
 DB 121 erfakacldnrnpeafqpmkgrglgaalkrshskgcncrsgclknyceyeaakimcss 180
 QY 181 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFGPPKLRKNROAFSCISW 240
 DB 181 icciacknyeesperkmlmstphymepgdfesshyllspakfsgppklrknroafscisw 240
 QY 241 EVVEATCACLQAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 DB 241 evveatcaccllaqgeaeqhcspslaemileefgrclsqilhieffkskglkme 295

RESULT 2
 AAY68464
 ID AAY68464 standard; Protein: 299 AA.
 AC AAY68464;
 XX
 XX 25-APR-2000 (first entry)
 DT Human testis specific factor tesmin SEQ ID NO:5.
 DE Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility.
 XX
 OS Homo sapiens.
 XX
 XX WO200004147-A1.
 XX
 XX 27-JAN-2000.
 XX
 XX 16-JUL-1999; 99WO-JP03859.
 XX
 XX 17-JUL-1998; 98JP-0219856.
 XX
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX
 XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX
 XX WPI; 2000-147785/13.
 XX N-PSDB; AAZ88156.
 XX
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 XX useful for investigation of germ cell differentiation and sterility -
 XX
 XX Claim 1; Page 50-52; 63pp; Japanese.

CC The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 299 AA;

Query Match 78.2%; Score 1251; DB 21; Length 299;
 Best Local Similarity 76.3%; Pred. No. 4.2e-111;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

QY 1 MVTCLKGGAGMLCIDNCCGARELKALHLLPOYDDQSFPQSELPKPMTTLVGRLLPVPK 60
 DB 1 mvtclkggagmlcidncgarelkahlhlpoyddqsfpqselpkpmttlvgrllvpak 60
 QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTITLGGYDCFSGDFCNCSNLRHEL 120
 DB 61 lnlitqvdngalpsavngaappspalqgpptitlsgydcfsgdfcnscnnlrhel 120
 QY 117 RHELEFRKAICLDNRNPEAFQPMKGRGLGAALKRHSKGCNCRSGCLKNYCEYEAQI 176
 DB 117 rhelefrkaicldnrnpeafqpmkgrglgaalkrshskgcncrsgclknyceyeaqi 180
 QY 177 MCSSICKIACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFGPPKLRKNROAFS 236
 DB 177 mcssickicgknyeesperkmlmstphymepgdfesshyllspakfsgppklrknroaf 240
 QY 237 CISWEVVEATCACLQAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 DB 241 ciswevveatcaccllaqgeaeqhcspslaemileefgrclsqilhieffkskglkme 299

RESULT 3
 ABB62035
 ID ABB62035 standard; Protein: 950 AA.
 XX
 AC ABB62035;
 XX
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster polypeptide SEQ ID NO 12897.
 DE Drosophila melanogaster polypeptide; cell signalling; insecticide;
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 XX
 XX 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX
 XX 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 XX N-PSDB; ABL06138.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -
 XX
 XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and

PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 DR N-PSDB; AAL01300.

XX Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a

CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a protein of the invention.
 XX Sequence 147 AA;
 SQ
 Query Match 20.1%; Score 322; DB 22; Length 147;
 Best Local Similarity 73.8%; Pred. No. 8e-23; Indels 0; Gaps 0;
 Matches 62; Conservative 7; Mismatches 15;
 Qy 1 MVICOLKGGQMLCIDNCGARELKALHLLPYDDQSSFPQSELPKPMTTLVGRLLPVPK 60
 Db 60 mvicqlkggtqmlcidnsrtrelkalhlvpqyqdgnylqsdvvpkmtalvgrflpstk 119
 Qy 61 LNLITQVDNGALPSAVNGAAPPSSG 84
 Db 120 lnlitqqlgalpsvngsaifpsg 143
 RESULT 7
 AAG76158
 ID AAG76158 standard; Protein; 53 AA.
 XX
 AC AAG76158;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE Human colon cancer antigen protein SEQ ID NO:6922.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 11.
 XX
 OS Homo sapiens.
 XX
 PN WO200122920-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26524.
 XX
 PR 29-SEP-1999; 99US-0157137.
 PR 03-NOV-1999; 99US-0163280.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;
 XX WPI; 2001-235357/24.
 DR N-PSDB; AAG76158.
 XX
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX Claim 11; Page 8365; 9803pp; English.
 CC
 CC AAG76158 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAG77788 to AAG7789
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

Qý 132 RNPEAFQPMKGRLGAALKRHSKGCNCKRRSGCLKNYCEYEAKIMCSSIPCIACKNYE 191
||: : |||: || : ||| ||| | : |||: ||: : ||: ||: ||:

QY 103 SSGDFCNSCENNLRHELERFKAIKACLDNRNPEAFQPKMGKGRGLGAAKLRRHSGKGCNCKRS 162

```

Db 1677 dcassggtltkcsyghnvtvfdsgsvtgrgcse-----svfeesdyc-denp-aycpr 1729
QY 141 MGKGRLLGAALKRHSKGCNCRSGCLKNYCEYEAKIMCSSTC-----KCI 185
Db 1730 c-----nngcntads--qdkyvecivcdsvdncvdsdptqitktqrqchercl 1776
QY 186 AC----KNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWE 241
Db 1777 safilplfnetedpsyalirncyddlekdred-----actagskrfcacatcegtknsed 1829
QY 242 VVENTACCLLAQGEA-----EQEHCSPSLAEQMILEEPGRCLSQILH 284
Db 1830 lvasrhscvlvcqgdccsqgasscnsyrehdecyigtdeersitslg-clseish 1883

RESULT 12
AAAY95660
ID AAY95660 standard; Protein; 602 AA.
XX
AC AAY95660;
XX
DT 25-OCT-2000 (first entry)
XX
DE Human Znt2 protein.
XX
KW Znt2; epidermal growth factor-like domain; human;
KW cell differentiation; vulnery; diagnosis; therapy.
XX
OS Homo sapiens.
XX
FH Key
FT Peptide
FT /label= Signal_peptide
FT /note= "the signal peptide alternatively
FT comprises amino acid residues 9..30"
FT
FT Protein
FT Domain
FT /note= "extracellular domain"
FT
FT Domain
FT /note= "transmembrane domain"
FT
FT Domain
FT /note= "intracellular domain"
FT
FT Region
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 224..256"
FT
FT Region
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 272..303"
FT
FT Region
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 317..351"
FT
FT Region
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 371..402"
FT
FT Region
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 422..449"
FT
FT WO200043512-A1.
XX
XX 27-JUL-2000.
XX
XX 20-JAN-2000; 2000WO-US01419.
XX
XX 25-JAN-1999; 99US-0237074.
XX
XX (ZYMO ) ZYMOGENETICS INC.
XX

```

```

PI Holloway JL, Lofton-Day CE, Gilbert T;
XX
DR WPI: 2000-491163/43.
DR N-PSDB; AAA50101, AAA50102.
XX
PT Isolated Znt2 nucleic acids and polypeptides which act as epidermal
PT growth factors, useful for the treatment of e.g. kidney and liver
PT disorders, burns, and ulcers and for regulating smooth muscle cell
PT proliferation.
XX
XX Claim 1; Page 89-91; 98pp; English.
XX
CC The present sequence is of that of human Znt2, a new member of the
CC epidermal growth factor-like domain group of proteins. The
CC sequence was deduced from isolated Znt2 cDNA (see AAA50101). In
CC addition to the presence sequence, polypeptides comprising amino
CC acids 31-507 and 31-602 of the sequence are claimed. An expression
CC vector, a recombinant host cell, a method of using the vector to
CC produce Znt2 protein, an antibody or antibody fragment that
CC specifically binds to the polypeptide, and a method of detecting
CC the presence of Znt2 gene products in a biological sample are also
CC all claimed. Znt2 polypeptides may be used to regulate vascular
CC smooth muscle cell proliferation, to restore normal neurological
CC functioning after trauma, to treat ocular disorders, to treat
CC kidney and liver disorders, to promote hair and follicular
CC development, to stimulate growth and differentiation of various
CC epidermal and epithelial cells in vivo and in vitro, for the
CC treatment of burns, ulcers and corneal incisions, and to stimulate
CC wound healing.
XX
SQ Sequence 602 AA;
XX
Query Match 6.1%; Score 97; DB 21; Length 602;
Best Local Similarity 21.8%; Pred. No. 1.5; Indels 74; Gaps 16;
Matches 56; Conservative 26; Mismatches 101;
QY 44 PKPMTTLVGRLLVPVPAKLNILITQVDNGALPSAVNGAAPPSPGALQGPPIKTLSCYCDIFS 103
Db 154 pap-ttpvattvpapt----tprtptdlpsnsnssvltptptateapspppeyvcnsv 208
QY 104 SGDF-----CN-----SCSCNNLRHELEFKAICACLDNRNPAFQPMKGKGRLL-----G 147
Db 209 vgslnvnrqnttgqccrpgyqglh----cetc-----kegfynytsgicqcdcsphg 260
QY 148 AAKL--RHSGKCNCK-----RSGLKNYCEYEAKIMCSICKCIA 186
Db 261 alsipcnssgkckvgvigsicdrqdggygfskngclp--ccnmrnsacdaltg--a 316
QY 187 CKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWEVVEAT 246
Db 317 clncqen-----skgnhce--eckegfygsp-----datkeclrc-pcsavtst 357
QY 247 CACLLAAGE-EAQGEHC 262
Db 358 gscsikssellepecdcq 374
XX
RESULT 13
ABB61083
ID ABB61083 standard; Protein; 1987 AA.
XX
AC ABB61083;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 10041.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX

```


PN W0200171042-A2.
 XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 PR (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX N-PSDB; ABL05186.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 10041; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
 CC sequences (AB101840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 1987 AA;
 SQ

Query Match 6.1%; Score 97; DB 22; Length 1987;
 Best Local Similarity 20.0%; Pred. No. 6.8;
 Matches 45; Conservative 20; Mismatches 82; Indels 78; Gaps 8;

Qy 65 TOVDNGALPSAV-----NGAARFSGPALQPPKAITLSGYCDGFCNSCNNLR 117
 Db 516 tkveaggtvsslllqlrqnsvnp--paistesqvl-----cakc---nlq 559
 Qy 118 HELERFKAIAKACLDNPEAFQPKMGKGRGAALRHSGKCNKSGCLNKVCCEYAKIM 177
 Db 560 fderferqeln-----khqacsgiqsandsqskqerffr 595
 Qy 178 CSSICK-----CIACKNVEESPERKMLMSTPHYMEPGDFESSHYLS 218
 Db 596 csqcsbvhaqwhffihmrevhrryiclycnhyvpsveklsl-----hlenkhdldqshfak 651
 Qy 219 PAKFSGPPKLRNROAFSCISWEVVEATCACLQAQGEAEQHS 263
 Db 652 dawdqqgskedarthlvcc-----tcqatfvgsfedhdcs 688

RESULT 14
 AAG09497
 ID AAG09497 standard; Protein; 359 AA.
 XX AAG09497;
 AC AAG09497;
 XX 17-OCT-2000 (first entry)
 DT Arabidopsis thaliana protein fragment SEQ ID NO: 7455.
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 7455.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.

XX EPI033405-A2.
 PN 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 PD 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 XX 09-MAR-1999; 99US-0123548.
 XX 23-MAR-1999; 99US-0125788.
 XX 25-MAR-1999; 99US-0126264.
 XX 29-MAR-1999; 99US-0126785.
 XX 01-APR-1999; 99US-0127462.
 XX 06-APR-1999; 99US-0128234.
 XX 08-APR-1999; 99US-0128714.
 XX 16-APR-1999; 99US-0129845.
 XX 19-APR-1999; 99US-0130077.
 XX 21-APR-1999; 99US-0130449.
 XX 23-APR-1999; 99US-0130510.
 XX 28-APR-1999; 99US-0130891.
 XX 30-APR-1999; 99US-0131449.
 XX 04-MAY-1999; 99US-0132407.
 XX 05-MAY-1999; 99US-0132484.
 XX 06-MAY-1999; 99US-0132485.
 XX 06-MAY-1999; 99US-0132486.
 XX 07-MAY-1999; 99US-0132487.
 XX 11-MAY-1999; 99US-0132863.
 XX 14-MAY-1999; 99US-0134256.
 XX 14-MAY-1999; 99US-0134218.
 XX 14-MAY-1999; 99US-0134219.
 XX 14-MAY-1999; 99US-0134221.
 XX 14-MAY-1999; 99US-0134370.
 XX 18-MAY-1999; 99US-0134768.
 XX 20-MAY-1999; 99US-0134941.
 XX 21-MAY-1999; 99US-0135124.
 XX 24-MAY-1999; 99US-0135353.
 XX 25-MAY-1999; 99US-0135629.
 XX 27-MAY-1999; 99US-0136021.
 XX 28-MAY-1999; 99US-0136392.
 XX 01-JUN-1999; 99US-0136782.
 XX 03-JUN-1999; 99US-0137222.
 XX 04-JUN-1999; 99US-0137528.
 XX 07-JUN-1999; 99US-0137502.
 XX 08-JUN-1999; 99US-0137724.
 XX 10-JUN-1999; 99US-0138094.
 XX 10-JUN-1999; 99US-0138540.
 XX 14-JUN-1999; 99US-0138847.
 XX 16-JUN-1999; 99US-0139119.
 XX 16-JUN-1999; 99US-0139452.
 XX 17-JUN-1999; 99US-0139453.
 XX 18-JUN-1999; 99US-0139454.
 XX 18-JUN-1999; 99US-0139455.
 XX 18-JUN-1999; 99US-0139456.
 XX 18-JUN-1999; 99US-0139457.
 XX 18-JUN-1999; 99US-0139458.
 XX 18-JUN-1999; 99US-0139459.
 XX 18-JUN-1999; 99US-0139460.
 XX 18-JUN-1999; 99US-0139461.
 XX 18-JUN-1999; 99US-0139462.
 XX 18-JUN-1999; 99US-0139463.
 XX 18-JUN-1999; 99US-0139730.
 XX 21-JUN-1999; 99US-0139763.
 XX 22-JUN-1999; 99US-0139817.
 XX 23-JUN-1999; 99US-0139899.
 XX 23-JUN-1999; 99US-0140353.
 XX 23-JUN-1999; 99US-0140354.
 XX 24-JUN-1999; 99US-0140695.
 XX 28-JUN-1999; 99US-0140823.
 XX 29-JUN-1999; 99US-0140991.
 XX 30-JUN-1999; 99US-0141287.

PR	16-SEP-1999;	99US-015403.9;
PR	20-SEP-1999;	99US-015477.9;
PR	22-SEP-1999;	99US-015513.9;
PR	23-SEP-1999;	99US-015548.6;
PR	24-SEP-1999;	99US-015556.9;
PR	28-SEP-1999;	99US-015645.8;
PR	29-SEP-1999;	99US-015656.6;
PR	04-OCT-1999;	99US-015711.7;
PR	05-OCT-1999;	99US-015757.3;
PR	06-OCT-1999;	99US-015786.5;
PR	07-OCT-1999;	99US-015802.2;
PR	08-OCT-1999;	99US-015823.2;
PR	12-OCT-1999;	99US-015836.9;
PR	13-OCT-1999;	99US-015929.3;
PR	13-OCT-1999;	99US-015929.4;
PR	13-OCT-1999;	99US-015929.5;
PR	14-OCT-1999;	99US-015932.9;
PR	14-OCT-1999;	99US-015933.0;
PR	14-OCT-1999;	99US-015933.1;
PR	14-OCT-1999;	99US-015963.7;
PR	14-OCT-1999;	99US-015963.8;
PR	18-OCT-1999;	99US-015958.4;
PR	21-OCT-1999;	99US-016074.1;
PR	21-OCT-1999;	99US-016076.7;
PR	21-OCT-1999;	99US-016076.8;
PR	21-OCT-1999;	99US-016077.0;
PR	21-OCT-1999;	99US-016081.4;
PR	21-OCT-1999;	99US-016081.5;
PR	22-OCT-1999;	99US-016098.0;
PR	22-OCT-1999;	99US-016098.1;
PR	22-OCT-1999;	99US-016098.9;
PR	25-OCT-1999;	99US-016140.4;
PR	25-OCT-1999;	99US-016140.5;
PR	25-OCT-1999;	99US-016140.6;
PR	26-OCT-1999;	99US-016135.9;
PR	26-OCT-1999;	99US-016136.0;
PR	26-OCT-1999;	99US-016136.1;
PR	28-OCT-1999;	99US-016192.0;
PR	28-OCT-1999;	99US-016199.2;
PR	28-OCT-1999;	99US-016199.3;
PR	29-OCT-1999;	99US-016214.2;
PR	29-OCT-1999;	99US-016214.3;

Query Match 0.0%; Score 90.5; DB 21; Length 339;
Best Local Similarity 32.4%; Pred. No. 0.84;
Matches : 34; Conservative 10; Mismatches 38; Indels 23; Gaps 7;

RESULT 15

AAG09496
 ID AAG09496 standard: Protein: 443 AA.

AC AAG09496;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 7454.

KW Protein identification; signal transduction pathway; metabolic pathway;

KW termination sequence.

OS *Arabidopsis thaliana*.

PN EP1033405-A2.



PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-0301439.
 XX XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143634.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 23-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.

```
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161320.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 6.0%; Score 96.5; DB 21; Length 443;
Best Local Similarity 32.4%; Pred. No. 1.1;
Matches 34; Conservative 10; Mismatches 38; Indels 23; Gaps 7;

Qy 131 DRNPEAFQPKMGKGRGAKLRHSGK---CNC-KRSGCLKNYCEYEAKIMCSSLICKCIA 186
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db 322 esnsepfqpetseetl--ckkeqskaeavccctkkslcktkckckangsgcgsgccla 379

Qy 187 --CKNYEES--PERKMLMSTPHYMEPGDFESSHYLSPAKFSGGPPK 227
: | | | | : | | | | : | | | | : | | | | : | | | | : | | | |
Db 380 skcsnreesakpdkp-----mepidvk-----kpagishddk 411
```

Search completed: July 11, 2002, 08:16:59
Job time: 203 sec

[illegible]

Query Match 7.28; Score 115; DB 2; Length 1737;

RESULT 7
T27827
hypothetical protein ZK287.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27827
R:MCMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z20425
A:Accession: T27827
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-497 <MIL>
A:Cross-references: EMBL:Z70757; PIDN:CAA94805.1; GSPDB:GN00023; CESP:ZK287.1
A:Experimental source: clone ZK287
C:Genetics:
A:Gene: CBSP:ZK287.1
A:Map position: 5
A:Introns: 20/1; 44/2; 99/3; 339/3; 443/2

Query Match 6.1%; Score 97; DB 2; Length 497;
Best Local Similarity 26.6%; Pred. No. 1.8;
Matches 37; Conservative 15; Mismatches 51; Indels 36; Gaps 10;

QY 82 PGGPALGPPKITLSGYDCFCSSGDFENS-CSCNNLRHELERFAIKACLDNRNEAFQPK 140
 |:| | | | | :||| :||| :||| :||| :||| :||| :||| :|||
Db 63 PAKPKNACKAK---SERCDCLRKEGSKLCACQI-----C--KKKEA--PK 103

QY 141 MGKRGLGAARK-RHSGKCNCCK--RSGCLKNYCEYEAKIMCSSICKCIA-CKNYEESPER 196
 ||:| :||| :||| :||| :||| :||| :||| :||| :|||
Db 104 K-----LAKVAPTSQCQCAKGKKQCVKKECACTVYGFCASCKCGDCTNGASK--- 154

QY 197 KMLMSTPHYMEPGDFESSH 215
 | | :| :|
Db 155 ---FSVPKHVQNGFLEHKH 170

RESULT 8
MMFFBI
laminin beta-1 chain precursor - fruit fly (*Drosophila melanogaster*)
N:Alternate names: laminin chain B1
C:Species: Drosophila melanogaster
C:date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C:Accession: A28783; S14462; B28783
R:Montell, D.J.; Goodman, C.S.
Cell 53, 463-473, 1988
A:title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals
A:Reference number: A28783; MUID:88210471
A:Accession: A28783
A:Molecule type: mRNA
A:Residues: 1-1790 <MONI>
A:Cross-references: EMBL:M19525
R:Montell, D.J.; Goodman, C.S.
submitted to the EMBL Data Library, June 1988
A:description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain r
A:Reference number: S14462
A:Molecule type: mRNA
A:Residues: 1-667,'L',669-725,'VT',728-947,950-1790 <MON2>
A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
C:Genetics:
A:Gene: lamB1
A:Cross-references: FlyBase:FBgn0002527
A:Map position: 2L 28D
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
C:function:
A:description: Interact with cells and with other basement membrane proteins to promo
C:Superfamily: laminin beta-1 chain; laminin-type EGf-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu
F:1-26/Domain; signal sequence #status predicted <SIG>

F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
 F:27-288/Domain: VI <DOM6>
 F:289-561/Domain: V <DOM5>
 F:290-354/Domain: laminin-type EGF-like homology <LE01>
 F:357-417/Domain: laminin-type EGF-like homology <LE02>
 F:420-477/Domain: laminin-type EGF-like homology <LE03>
 F:480-528/Domain: laminin-type EGF-like homology <LE04>
 F:531-561/Domain: laminin-type EGF-like homology <LE05>
 F:562-789/Domain: IV <DOM4>
 F:643-645/Region: cell attachment (R-G-D) motif
 F:790-1189/Domain: III <DOM3>
 F:791-836/Domain: laminin-type EGF-like homology <LE06>
 F:839-882/Domain: laminin-type EGF-like homology <LE07>
 F:885-932/Domain: laminin-type EGF-like homology <LE08>
 F:935-990/Domain: laminin-type EGF-like homology <LE09>
 F:968-972/Region: cell adhesion #status predicted
 F:993-1042/Domain: laminin-type EGF-like homology <LE10>
 F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
 F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
 F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
 F:1190-1407/Domain: II <DOM2>
 F:1408-1434/Domain: alpha <ALP>
 F:1435-1790/Domain: I <DOM1>
 F:51-56/Disulfide bonds: #status predicted
 F:140-203,234,489,593,1053,1248,1303,1332,1343,1475,1495,1517,1583,1646,1705/Binding site
 F:1191,1194,1788/Disulfide bonds: interchain #status predicted

Query Match 6.0%; Score 96.5; DB 1; Length 1790;
 Best Local Similarity 21.0%; Pred. No. 7.2;
 Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;
 QY 94 TLSGVCDFSGDF-----CNSCSG-----NNLRHELERFKAIAKALDRNPEAFQPKM 141
 DB 1021 TTGDHCELCRKGDFGDALQNCQCECDFLGTNTIAHCDRFTQCPCCL-----PNV 1072
 QY 142 GKRLGAKLHRSK-----GNCRRSGCLKNYCEYPAKIMCSSICKIACK-----N 189
 DB 1073 QGVRCQCAENHWKIASGEGSCNCDPIGALHEQCNSYTGQCCKPFGGRACNQCAH 1132
 QY 190 YEESPERKMLMSTHYMEPGDFESSHYLSPAKFSPPKLRNQAFAFSCISWEVVEATCAC 249
 DB 1133 YWGNPNKQPCQCDQGAADFQC-----DRTGNCVCHGIGGYKCN 1175
 QY 250 LLAQGEAEQHCSP 264
 DB 1176 ECARGYIQPFHCSP 1190
 RESULT 9
 S55554
 male-specific lethal 2 protein - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
 C:Accession: S55554
 R:Zhou, S.; Yang, Y.; Scott, M.J.; Pannuti, A.; Fehr, K.C.; Eisen, A.; Koonin, E.V.; Fod
 EMBO J. 14, 2884-2895, 1995
 A:Title: Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex
 A:Reference number: S55554; MUID:95317307
 A:Accession: S55554
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-769 <ZHO>
 A:Cross-references: EMBL:Z48443; NID:g872110; PIDN:CAA88358.1; PID:g872111
 C:Genetics:
 A:Gene: FlyBase:msl-2
 A:Cross-references: FlyBase:FBgn0005616
 A:Introns: 641/1
 Query Match 6.0%; Score 96; DB 2; Length 769;
 Best Local Similarity 25.7%; Pred. No. 3.3;
 Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;

QY 121 ERFKAIAKAC-----LDRNPEAFQPKMGKGRGLGAALRHSKG-----CNCKRS 163
 DB 471 KRFTLKASQAAKIEPVSEVKTKVQSGK-GA--LRRIKGRKEKVKPKPKRCGSG 527
 QY 164 -----CLANYCECYEAKIMCSSICKIACKN-----YEES-----PER 196
 DB 528 SSNTLTTCRNSRCPCYKSYNSCAG-CHCVCKNPKHEDYVESEDDDDLEDFEMPKDVP 586
 QY 197 KMLMSTHYMEPGDFESSHYLSPAKFSPPK 228
 DB 587 MTQSEEPVVAEPQRENS--MAPDSSAPISL 616
 RESULT 10
 T10053
 laminin alpha 5 chain - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
 C:Accession: T10053
 R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
 submitted to the EMBL Data Library, November 1997
 A:Reference number: Z16923
 A:Accession: T10053
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3635 <MIN>
 A:Cross-references: EMBL:U37501; NID:g2599231; PID:g2599232
 C:Genetics:
 A:Gene: Lama5
 C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like h
 C:Key words: basement membrane; cell binding; extracellular matrix
 F:1888-1939/Domain: laminin-type EGF-like homology <LEG>
 F:1942-1970/Domain: EGF homology <EGF>

Query Match 6.0%; Score 96; DB 2; Length 3635;
 Best Local Similarity 18.0%; Pred. No. 16;
 Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;
 QY 14 CIDNCGARELKALHLLPOYDQSSFSFQSELPKPMTWLLVGLLPVPAKLNLITVDN 73
 DB 117 CLERFGERTLER---ITQDDVI-----CTTEYSRIVPL-ENGEIIVSLVNGR-P 161
 QY 74 SAVNGAAPP-----SGPALQCP-----PKITLSGY 98
 DB 162 GALNFSYSPLLRDFTKATNIRLRLTNTLLGLMKGKALRDPVTRYYYSIKDISIGR 221
 QY 99 CDCFSSGDFC-----NSC-----SC 113
 DB 222 CVCHGHADVCDKADPLDPFRLQCAQHNTCGSCDRCCPGFNQOPWKPATPDTSDANE 281
 QY 114 NNLRH-----ELERFKAIAKALDRNPEAFQPKMGKGRGLGAALRHSKCNCKRS 162
 DB 282 NCHGHAYDCYVDPEVDR-----RNASQNDNVITQG--GGVCLDCQHHTTGINGER- 329
 QY 163 GCLKNY-----CECYEAKI---MCSSI-----CK-----CIAC-KNY 190
 DB 330 -CLPGFRAPDQPLDSPHVCRPCDC-ESDFTDGTCDLTGRCYCVRPNFTCELCAACAE 387
 QY 191 EESPERKMLMSTHYMEPGDFESSHYLSPAKFSPPKLRNQAFAFSCISWEVVEATCAC 250
 DB 388 TDFPHCYPLPSFPHN-----DTRQVILPA-----QOIVNCDNAA 422
 QY 251 LAQGEAEQ-----HC-----SPSLAEQMTLEEFGR 277
 DB 423 GTQGNACRKDRLGRVCVKPNFRGAHCELCAPGFHGSPCHPCCSPGVSANSLCDPESG 482
 QY 278 CL 279
 DB 483 CM 484

RESULT 11

S61917
protein kinase C (EC 2.7.1.1-) PKCA - Aspergillus niger
C:Species: Aspergillus niger
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S61917
R:Morawetz, R.; Lendenfeld, T.; Mischak, H.; Muehlbauer, M.; Gruber, F.; Goodnight, J.;
Mol. Gen. Genet. 250, 17-28, 1996
A:Title: Cloning and characterization of genes (pkc1 and pkcA) encoding protein kinase C
A:Reference number: S61917; MUID:96158841
A:Accession: S61917
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1096 <MOR>
A:Cross-references: EMBL:U10549; NID:g507899; PIDN:AAA97433.1; PID:g507900
C:Genetics:
A:Gene: pkcA
A:introns: 145/1; 234/1; 428/1; 560/1; 698/1; 874/2; 941/3; 1077/1
C:Superfamily: yeast protein kinase C; protein kinase C zinc-binding repeat homology; PK
C:Keywords: ATP; duplication; phospholipid binding; phosphotransferase; serine/threonine
F:460-507/Domain: protein kinase C zinc-binding repeat homology <K21>
F:528-577/Domain: protein kinase C zinc-binding repeat homology <K22>
F:769-1030/Domain: protein kinase homology <KIN>
F:777-785/Region: protein kinase ATP-binding motif

Query Match

Best Local Similarity 5.9%; Score 95; DB 1; Length 1096;

Matches 63; Conservative 21; Mismatches 91; Indels 104; Gaps 15;

QY 66 QVNDGALPSAVN-----GAFFSPGALOGPKKITLSGYCDGFCSSGDFCNCSNNLRHEL 120

Db 359 KMEHGAGPRDPTAGGSAHPSGPG--GPGAAPSGQYGGSGAGAPAGNQVMDSW-FAL 415

QY 121 ERFKAI-----KACLDNRPAPQPKMGKGLGAAKLR-----HSKG-----C 157

Db 416 EPVGRILYSMSFAKQLKDRPT-----DGLNRQGAVRQKEEVEHKGKHFVYQOYNIM 471

QY 158 NCKRSGLKNYC---ECYEAKTWCSSIC-----KCIACKNYEESPERKMLMSTPHMEP 208

Db 472 RALCGDFLKYANGQACDKYTKCHKYKPKVYTKISKANETDP-----517

QY 209 GDFESSHYLSPAKFSGPPKLRKRNQAFSCISWEVVEATCACLGAOGEAEQEHCSPLAE 268

Db 518 -DEEKINHRIPHFEG-----FSNIS-----ANWCC-----HCG-----545

QY 269 QMILEFGR-----CLSQILHIEFKSKGLKIE 295

Db 546 --YLLPFGKRNKARCTECGLTCHSHCTHLVDPFCGMSME 582

RESULT 12

T08852
lustrin A - California red abalone
C:Species: Haliotis rufescens (California red abalone)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08852
R:Shen, X.; Belcher, A.M.; Hansma, P.K.; Stucky, G.D.; Morse, D.E.
J. Biol. Chem. 272, 32472-32481, 1997
A:Title: Molecule cloning and characterization of lustrin A, a matrix protein from shell
A:Reference number: Z16496; MUID:98070424
A:Accession: T08852
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1428 <SHE>
A:Cross-references: EMBL:AF023459; NID:g2723361; PIDN:AAB95154.1; PID:g2723362
A:Experimental source: tissue type mantle (shell and pearl nacre); cell type pallial
C:Superfamily: antileukoproteinas repeat homology
C:Keywords: extracellular matrix; extracellular protein
F:1382-1426/Domain: antileukoproteinas repeat homology <ALP>

Query Match

Best Local Similarity 5.9%; Score 95; DB 2; Length 1428;

Best Local Similarity 22.8%; Pred. No. 7.6;
Matches 56; Conservative 21; Mismatches 81; Indels 88; Gaps 14;

QY 74 SAVNGAAPPSPGAL-----OGPKKITLSGYCDGFCSS-----GDFCN-----S 110

Db 409 SCVGAACPRNTACFAAPSGSPAV-----C-CYTSPPRPPEPPSPSPPTGDCPCGPNVN 462

QY 111 CSCNNLR-----HELERFKAIAKACLDNRPAPQ-----KMGKRLGAALKRHSKGCNCKRS 162

Db 463 CTAGTCRLVEDCRIRCCPAVPKCIDRPPPLPPDPVCPVGLGA-----DLKLYCGRG 518

QY 163 G-CLKN-YCECYEA-----KIMCS-SICKCIA 186

Db 519 GRRCPWNTYCVIHPADYAVCCFGSPASAIAPTSAPGVPDPCPCGVNVNCTIGVCRLVA 578

QY 187 CANYEESPERKMLMSTPHMEFGDFESSHYLSPAKFSGPPKLRKRNQAFSCISWEVVEAT 246

Db 579 NCDYPCPARPTCVD--HSSEP-----SLNCTIGDPALNGKLEEFSCVGGRLCPLN 627

QY 247 CACLLA 252

Db 628 TACLAA 633

RESULT 13

T26998

hypothetical protein Y48B6A.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26998

R:Wall, M.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z20297

A:Accession: T26998

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-716 <WIL>

A:Cross-references: EMBL:AL110490; NID:el542263; PIDN:CAB54442.1; CESP:Y48B6A.6

A:Experimental source: clone Y48B6A

C:Genetics:

A:Gene: CESP:Y48B6A.6

A:introns: 38/3; 196/3; 437/2; 460/3; 518/1; 617/3; 673/3

Query Match

Best Local Similarity 25.0%; Pred. No. 4.6;

Matches 58; Conservative 33; Mismatches 95; Indels 46; Gaps 10;

QY 22 ELKALHLLPOYDQ-----SSFPQSELPKPMTTLVGRL---LPVPAKLNLIQVD- 68

Db 378 EVKALEAQASDDRPASPTPIRDSSLPPLPPPKPETPLAIRAGPIPTQLEMIHQEDC 437

QY 69 --NGALPSAVNGAAPPSPGALOGPKKITLSGYCDGFCSSGDFCNCSNNLRHELERFKAI 126

Db 438 SIRSPSPTSVSHGSRQSPAPVKPKSVTVS-----PLGLLCDP-----NL--SIEKPEEM 485

QY 127 KACLDNRPAPQPKMGKRLGAALKRHSKGCNCKRSGLKNYCECYEAKIMCSSICKCIA 186

Db 486 KTEDTKPVETAPAPVDEAEALNDALDRNRK---INEATCL-----TKIFFSLVSKYFQ 534

QY 187 CANYEESPERKMLMSTPHMEFGDFESSHYLSPAKFSGPPKLRKRNQAFSCI 238

Db 535 NCSFDFLNKRSGENVPLKMK---KISLYAEFESEFS-----RKQIQYFSGI 578

RESULT 14

A71928

cag island protein - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: A71928

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.

A:Description: Isolation of a partial cDNA encoding a protein homologous to laminin A
A:Reference number: S40150
A:Accession: S40150
A:Molecule type: mRNA
A:Residues: 1403-1541,'S',1543-1816 <RT3>
A:Cross-references: EMBL:X70904; NID:g437804; PIDN:CAA50261.1; PID:g437805
C:Genetics:
A:Gene: GDB:LAMA4; LAMA3
A:Cross-references: GDB:203904; OMIM:600133
A:Map position: 6q21-6q21
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
C:Function:

C;Superfamily: laminin alpha-4 chain; laminin G repeat homology; laminin type IGF
C;Keywords: basement membrane; cell binding; collagen; extracellular matrix; glyco
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1816/Product: laminin alpha-4 chain #status predicted <NAT>

F; 132-184/Domain: laminin-type EGF-like homology <LE2>
F; 132-184/Domain: laminin-type EGF-like homology <LE2>
F; 187-238/Domain: laminin-type EGF-like homology <LE3>
F; 241-265/Domain: laminin-type EGF-like homology #status atypical <LE4>

F; 862-1031/Domain: laminin G repeat homology <LG1>
F; 1068-1223/Domain: laminin G repeat homology <LG2>
F; 1252-1398/Domain: laminin G repeat homology <LG3>

F:1665-1816/Domain: laminin G repeat homology <IG5>
F:104,215,308,458,524,550,571,574,631,639,735,751,754,780,803,1086,1281,1359,1411/Bin
F:266,269/Dissulfide bonds: interchain status predicted

Query Match 5.8%; Score 93; DB 1; Length 1816;
Best Local Similarity 23.8%; Pred. NO. 14;

QY 2 VTCOLKGAQMLCIDNCG-----ARELKALH---LLPQY--DDQSFQSQSELPKPMWTLV 51

QY 52 GRLLPVPKLNLTQV--DNGALPSAVN-GAAFPS-----GPAHQGPCKITLSGYCDFSS 104

QY 105 GDFCNCSGN-----NLRHE--LEREFAIKACLDNRPEAFQ-PKMGKGRLCAAKL-RHSK 155

181 GTGCGAGCCGCSCSNRIITFCDCDYMCOCBNCI-BMTCBCEBECACVVCDBAIAKNA 239

156 GCNCRSGCLKNYCEYEAKIMCSIC-KCI-----ACKNYEESPCKMLMST--P 203 QY
240 VANGCEGCDGCVMTGAI EGGCEBDETCGCVGVNFI TDDI PYNISTEECKAVI SVSSGAA 288 QY
240 VANGCEGCDGCVMTGAI EGGCEBDETCGCVGVNFI TDDI PYNISTEECKAVI SVSSGAA 288 QY

204 QY H Y M E P G D F E S S H Y L S P A K F S G P P K L R K N R Q A F S C I S W E V E A T C A C I L L A Q E A E A Q E H C S 263

QY 264 PSLAEOMILEE 274
1 : : |
255 ADDCALLEY 255

	C
	S
	V
	T
	A
	C
	C
	S
	T
	:
	F
	T
	4
	(
	r
	E
	(
	J
	(
)
	(

Job time: 104 sec

Search completed: July 11, 2002, 08:15:20
Job time: 104 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:15:26 ; Search time 30.23 seconds
(without alignments)
377.845 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRLCSQILHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1599	100.0	295	1 MTL5_MOUSE	Q9wtj6 mus musculus
2	1251	78.2	299	1 MTL5_HUMAN	Q9y415 homo sapien
3	96.5	6.0	1790	1 LMB1_DROME	P11046 drosophila
4	96	6.0	773	1 MSL2_DROME	P50534 drosophila
5	96	6.0	3718	1 LMA5_MOUSE	Q61001 mus musculus
6	95	5.9	1096	1 KPC1_ASPNG	Q00078 aspergillus
7	93	5.8	1816	1 LMA4_HUMAN	Q16363 homo sapien
8	92	5.8	798	1 ITB1_XENLA	P12606 xenopus lae
9	91	5.7	493	1 OC90_HUMAN	Q02509 homo sapien
10	91	5.7	747	1 E2H1_HUMAN	Q92800 homo sapien
11	91	5.7	2907	1 FBN2_MOUSE	Q61555 mus musculus
12	89	5.6	1786	1 LMB1_MOUSE	P02469 mus musculus
13	88	5.5	798	1 ITB0_XENLA	P12607 xenopus lae
14	88	5.5	1786	1 LMB1_HUMAN	P07942 homo sapien
15	87	5.4	290	1 Y070_NPVAC	Q11470 autographa
16	87	5.4	501	1 TRA2_HUMAN	Q12933 homo sapien
17	87	5.4	652	1 SVG_SCHPO	Q10179 schizosacch
18	86.5	5.4	611	1 LEM2_CANFA	P33730 canis famil
19	86	5.4	610	1 LEM2_HUMAN	P16581 homo sapien
20	86	5.4	747	1 E2H1_MOUSE	P70351 mus musculus
21	86	5.4	830	1 LEM3_HUMAN	P16109 homo sapien
22	85.5	5.3	769	1 ITB7_HUMAN	P05107 homo sapien
23	85.5	5.3	806	1 ITB7_MOUSE	P26011 mus musculus
24	85	5.3	398	1 MUB1_XENLA	P38565 xenopus lae
25	85	5.3	746	1 E2H2_MOUSE	Q61188 mus musculus
26	85	5.3	787	1 ITB6_MOUSE	Q9z0t9 mus musculus
27	84.5	5.3	169	1 KRGA_HUMAN	P26371 homo sapien
28	84.5	5.3	720	1 NRG3_HUMAN	P56975 homo sapien
29	84.5	5.3	1639	1 LMCL_DROME	P15215 drosophila
30	84.5	5.3	1816	1 LMA4_MOUSE	P97927 mus musculus
31	84	5.3	769	1 ITB2_BOVIN	P32592 bos taurus
32	84	5.3	2531	1 NTC1_MOUSE	Q01705 mus musculus
33	84	5.3	2531	1 NTC1_RAT	Q07008 rattus norv

34	84	5.3	3707	1 PGBM_MOUSE	Q05793 mus musculus
35	84	5.3	3828	1 TRX_DROVI	Q24742 drosophila
36	83.5	5.2	419	1 VEGC_HUMAN	P49767 homo sapien
37	83.5	5.2	769	1 ITB2_PIG	P53714 sus scrofa
38	83.5	5.2	798	1 ITB7_HUMAN	P26010 homo sapien
39	83.5	5.2	2766	1 THYG_MOUSE	O08710 mus musculus
40	83	5.2	423	1 MKR2_SEROU	Q9dd48 seriola qui
41	83	5.2	746	1 E2H2_HUMAN	Q15910 homo sapien
42	83	5.2	771	1 ITB2_MOUSE	P11835 mus musculus
43	83	5.2	1877	1 PKK5_MOUSE	Q04592 mus musculus
44	82.5	5.2	445	1 AD18_RAT	P97776 rattus norv
45	82.5	5.2	768	1 LEM3_MOUSE	Q01102 mus musculus

ALIGNMENTS

RESULT 1	
MTL5_MOUSE	
ID MTL5_MOUSE	STANDARD; PRT; 295 AA.
AC Q9WTJ6;	
DT 16-OCT-2001 (Rel. 40, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific	
DE metallothionein-like protein).	
GN MTL5.	
OS Mus musculus (Mouse).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	
OX NCBI_TaxID=10090;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=ICR; TISSUE=Testis;	
RX MEDLINE=99208669; PubMed=10191092;	
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;	
RT "A novel testis-specific metallothionein-like protein, tesmin, is an	
RT early marker of male germ cell differentiation.";	
RL Genomics 57:130-136(1999).	
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.	
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.	
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES	
CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED	
CC PROGRESSIVELY.	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
DR EMBL; U77383; AAD24667.1; -.	
DR EMBL; U61716; AAD24666.1; -.	
DR MGD; MGI:1340029; Mtl5.	
KW Spermatogenesis.	
FT DOMAIN 99 187 CYS-RICH.	
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;	

Query Match 100.0%; Score 1599; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 8.2e-132;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1	MVICQLKGAQMLCIDNCGARELKALHLLPQYDDQSSFFQSQSELPKPMPTTLVGRLLPVPK 60
Db	1 MVICQLKGAQMLCIDNCGARELKALHLLPQYDDQSSFFQSQSELPKPMPTTLVGRLLPVPK 60
QY 61	LNLIQTVDNGALPSAVNGAAPPSPGALQGPKKITLSGVCDCPSSGDFCNSCSCNNLRHEL 120
Db	61 LNLIQTVDNGALPSAVNGAAPPSPGALQGPKKITLSGVCDCPSSGDFCNSCSCNNLRHEL 120

QY 121 ERKATKACLDNRNPEAFQPKMGKGRGAAKLRHSGKGCNCRSGCLKNYCECYEAKIMCSS 180
DB 121 ERKATKACLDNRNPEAFQPKMGKGRGAAKLRHSGKGCNCRSGCLKNYCECYEAKIMCSS 180
QY 181 IKCIACKNYEESPERKMLMSPHYMEPGDFESSHYLSYPAKSGPPKLRNROAFSCISW 240
DB 181 IKCIACKNYEESPERKMLMSPHYMEPGDFESSHYLSYPAKSGPPKLRNROAFSCISW 240
QY 241 EVVEATCACLQAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
DB 241 EVVEATCACLQAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
RESULT 2
MTLS_HUMAN
ID MTLS_HUMAN STANDARD; PRT; 299 AA.
AC Q9Y415;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific
metalothionein-like protein).
GN MTLS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, testin, is an
RL early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U86074; AAD24668.1; -
DR MIM: 604374; -
KW Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH.
SQ SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;
Query Match 78.2%; Score 1251; DB 1; Length 299;
Best Local Similarity 76.3%; Pred. No. 1.4e-101;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;
QY 1 MVICOLGGAQMCIDNCGARELKALHLLPOYDDQSFPOSELPKMTTLVGRLLPVPK 60
DB 1 MVICOLGGAQMCIDNCGARELKALHLLPOYDDQSFPOSELPKMTTLVGRLLPVPK 60
QY 61 LNLITQVNDGALPSAVNGAAPPSPGALOGPPKIFLGSYDCGFCSSGDCNCSG----CNLL 116
DB 61 LNLITQVNDGALPSAVNGAAPPSPGALOGPPKIFLGSYDCGFCSSGDCNCSG----CNLL 120
QY 117 RHLEERKAIKACLDNRNPEAFQPKMGKGRGAAKLRHSGKGCNCRSGCLKNYCECYEAKI 176
DB 117 RHLEERKAIKACLDNRNPEAFQPKMGKGRGAAKLRHSGKGCNCRSGCLKNYCECYEAKI 180
QY 177 MCSSTICKIACKNYEESPERKMLMSPHYMEPGDFESSHYLSYPAKSGPPKLRNROAFS 236
DB 177 MCSSTICKIACKNYEESPERKMLMSPHYMEPGDFESSHYLSYPAKSGPPKLRNROAFS 240
QY 181 MCSSTICKIACKNYEESPERKMLMSPHYMEPGDFESSHYLSYPAKSGPPKLRNROAFS 240
DB 181 MCSSTICKIACKNYEESPERKMLMSPHYMEPGDFESSHYLSYPAKSGPPKLRNROAFS 240

QY 237 CISWEVVEATCACLQAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
DB 241 CISWEVVEATCACLQAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 299
RESULT 3
LMBL_DROME
ID LMBL_DROME STANDARD; PRT; 1790 AA.
AC P11046; Q26328; Q9X2T4; Q9VLM6;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LMBL OR LMBL OR CG7123.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=94000382; PubMed=8397815;
RA Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;
RT "Analysis of the Drosophila gene for the laminin B1 chain.";
RL DNA Cell Biol. 12:573-587(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88210471; PubMed=3365769;
RA Montell D.J., Goodman C.S.;
RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain
RT reveals domains of homology with mouse.";
RL Cell 53:463-473(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C.R., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Berman B.P., Bhandari D., Bolshakov S.,
RA Beeson K.Y., Benos P.V., Brokstein P., Brokstein P., Brokstein P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of Drosophila melanogaster.";

CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.

CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; M95811; AAD19752.1; -.

DR EMBL; M19525; AAA28663.1; -.

DR EMBL; AE003618; AAF52563.1; -.

DR PIR; A28783; MMEFBI.

DR HSSP; P02468; TITLE.

DR FlyBase; FBgn0002527; LanBL.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001886; LamNT.

DR InterPro; IPR002049; Laminin_EGF.

DR Pfam; PF00053; laminin_EGF; 13.

DR Pfam; PF00055; laminin_Nterm; 1.

DR PRINTS; PR00011; EGF_LAMININ.

DR ProDom; PD002082; LamNT; 1.

DR SMART; SM00180; EGF_Lam; 11.

DR SMART; SM00001; EGF_Like; 1.

DR SMART; SM00136; LamNT; 1.

DR PROSITE; PS00022; EGF_1; 10.

DR PROSITE; PS01186; EGF_2; 2.

DR PROSITE; PS01248; LAMININ_TYPE_EGF; 12.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

FT SIGNAL 1 25

FT CHAIN 26 1790 LAMININ BETA-1 CHAIN.

FT DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).

FT DOMAIN 290 356 LAMININ EGF-LIKE 1.

FT DOMAIN 357 419 LAMININ EGF-LIKE 2.

FT DOMAIN 420 479 LAMININ EGF-LIKE 3.

FT DOMAIN 480 530 LAMININ EGF-LIKE 4.

FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).

FT DOMAIN 562 789 LAMININ DOMAIN IV.

FT DOMAIN 791 838 LAMININ EGF-LIKE 6.

FT DOMAIN 839 884 LAMININ EGF-LIKE 7.

FT DOMAIN 885 934 LAMININ EGF-LIKE 8.

FT DOMAIN 935 992 LAMININ EGF-LIKE 9.

FT DOMAIN 993 1044 LAMININ EGF-LIKE 10.

FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.

FT DOMAIN 1096 1143 LAMININ EGF-LIKE 12.

FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.

FT DOMAIN 1191 1407 DOMAIN II.

FT DOMAIN 1408 1434 DOMAIN ALPHA.

FT DOMAIN 1435 1790 DOMAIN I.

FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).

FT DOMAIN 1455 1507 COILED COIL (POTENTIAL).

FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).

FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).

FT	DISULFID	290	BY SIMILARITY.
FT	DISULFID	292	BY SIMILARITY.
FT	DISULFID	322	BY SIMILARITY.
FT	DISULFID	334	BY SIMILARITY.
FT	DISULFID	357	BY SIMILARITY.
FT	DISULFID	359	BY SIMILARITY.
FT	DISULFID	387	BY SIMILARITY.
FT	DISULFID	396	BY SIMILARITY.
FT	DISULFID	399	BY SIMILARITY.
FT	DISULFID	420	BY SIMILARITY.
FT	DISULFID	422	BY SIMILARITY.
FT	DISULFID	448	BY SIMILARITY.
FT	DISULFID	450	BY SIMILARITY.
FT	DISULFID	462	BY SIMILARITY.
FT	DISULFID	477	BY SIMILARITY.
FT	DISULFID	480	BY SIMILARITY.
FT	DISULFID	482	BY SIMILARITY.
FT	DISULFID	502	BY SIMILARITY.
FT	DISULFID	511	BY SIMILARITY.
FT	DISULFID	514	BY SIMILARITY.
FT	DISULFID	528	BY SIMILARITY.
FT	DISULFID	791	BY SIMILARITY.
FT	DISULFID	793	BY SIMILARITY.
FT	DISULFID	810	BY SIMILARITY.
FT	DISULFID	812	BY SIMILARITY.
FT	DISULFID	824	BY SIMILARITY.
FT	DISULFID	839	BY SIMILARITY.
FT	DISULFID	841	BY SIMILARITY.
FT	DISULFID	860	BY SIMILARITY.
FT	DISULFID	872	BY SIMILARITY.
FT	DISULFID	885	BY SIMILARITY.
FT	DISULFID	887	BY SIMILARITY.
FT	DISULFID	901	BY SIMILARITY.
FT	DISULFID	904	BY SIMILARITY.
FT	DISULFID	916	BY SIMILARITY.
FT	DISULFID	932	BY SIMILARITY.
FT	DISULFID	935	BY SIMILARITY.
FT	DISULFID	937	BY SIMILARITY.
FT	DISULFID	964	BY SIMILARITY.
FT	DISULFID	976	BY SIMILARITY.
FT	DISULFID	993	BY SIMILARITY.
FT	DISULFID	1007	BY SIMILARITY.
FT	DISULFID	1014	BY SIMILARITY.
FT	DISULFID	1017	BY SIMILARITY.
FT	DISULFID	1026	BY SIMILARITY.
FT	DISULFID	1029	BY SIMILARITY.
FT	DISULFID	1045	BY SIMILARITY.
FT	DISULFID	1047	BY SIMILARITY.
FT	DISULFID	1066	BY SIMILARITY.
FT	DISULFID	1068	BY SIMILARITY.
FT	DISULFID	1080	BY SIMILARITY.
FT	DISULFID	1093	BY SIMILARITY.
FT	DISULFID	1096	BY SIMILARITY.
FT	DISULFID	1098	BY SIMILARITY.
FT	DISULFID	1117	BY SIMILARITY.
FT	DISULFID	1129	BY SIMILARITY.
FT	DISULFID	1141	BY SIMILARITY.
FT	DISULFID	1144	BY SIMILARITY.
FT	DISULFID	1146	BY SIMILARITY.
FT	DISULFID	1163	BY SIMILARITY.
FT	DISULFID	1165	BY SIMILARITY.
FT	DISULFID	1177	BY SIMILARITY.
FT	DISULFID	1191	BY SIMILARITY.
FT	DISULFID	1194	BY SIMILARITY.
FT	DISULFID	1194	INTERCHAIN (PROBABLE).
FT	DISULFID	1788	INTERCHAIN (PROBABLE).
FT	SITE	643	CELL ATTACHMENT SITE (POTENTIAL).
FT	CARBOHYD	140	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	203	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	489	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	593	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1053	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1248	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1303	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1332	N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 6.0%; Score 96.5; DB 1; Length 1790;

Best Local Similarity 21.0%; Pred. No. 1.9;

Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;

QY 94 TSLGYDCDFSSGDF-----CNSCS-----NLRHELEKFAIKACLDNRNPEAFQPKM 141

DB 1021 TTGDHCELCCKGFFDALQQNCQCECDPLGTNTNTIAHCDRTFGQCPCL-----PNV 1072

QY 142 GKGRIGAAKLRHSK-----GCNCKRSGLKNYCEYEAKIMCSICKACK-----N 189

DB 1073 QGVRCDOCAENHWKIASGEGCSCNCDPIGALHEQCNSYTGQCKPGFGGRACNOCQAH 1132

QY 190 YEESPERKMLMSTPHYMPCGDFESSHYLSPAKFSGPPKLRKNQAFSCISWEVVEATCAC 249
 DR PROSITE; PS00518; ZF_RING_1; 1.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 KW Zinc-finger; Nuclear protein; DNA-binding; Coiled coil.
 FT ZN_FING 41 85 RING-TYPE
 FT DOMAIN 283 519 COILED COIL (POTENTIAL).
 FT CONFLICT 280 280 T -> N (IN REF. 2).
 FT CONFLICT 370 373 MISSING (IN REF. 2).
 FT CONFLICT 371 380 LKT -> VEE (IN REF. 3).
 FT CONFLICT 380 380 P -> Q (IN REF. 2).
 FT CONFLICT 403 403 T -> I (IN REF. 2).
 FT CONFLICT 403 403 T -> A (IN REF. 2).
 FT CONFLICT 405 405 T -> A (IN REF. 2).
 FT CONFLICT 423 423 H -> Q (IN REF. 3).
 FT CONFLICT 429 429 Q -> H (IN REF. 2).
 FT CONFLICT 431 431 A -> E (IN REF. 2 AND 3).
 FT CONFLICT 433 433 S -> P (IN REF. 2 AND 3).
 FT CONFLICT 439 439 E -> V (IN REF. 2).
 FT CONFLICT 443 443 A -> P (IN REF. 2).
 FT CONFLICT 471 471 T -> A (IN REF. 3).
 FT CONFLICT 591 591 M -> L (IN REF. 3).
 FT CONFLICT 642 642 E -> D (IN REF. 2).
 SQ SEQUENCE 773 AA; 84841 MW; AE38BA76A4174225 CRC64;

RESULT 4
 MSL2_DROME STANDARD; PRT; 773 AA.
 AC P50534;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Male-specific lethal-2 protein.
 GN MSL-2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95300219; PubMed=7781064;
 RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
 RA Kuroda M.I.;
 RT "Expression of msl-2 causes assembly of dosage compensation
 RT regulators on the X chromosomes and female lethality in Drosophila.";
 RL Cell 81:867-877(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95317307; PubMed=7796814;
 RA Zhou S., Yang Y., Scott M.J., Pannuti A., Fehr K.C., Eisen A.,
 RA Koonin E.V., Fouts D.L., Wrightman R., Manning J.E., Lucchesi J.C.;
 RT "Male-specific lethal 2, a dosage compensation gene of Drosophila,
 RT undergoes sex-specific regulation and encodes a protein with a RING
 RT finger and a metallothionein-like cysteine cluster.";
 RL EMBO J. 14:2884-2895(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96017637; PubMed=7588059;
 RA Bashaw G.J., Baker B.S.;
 RT "The msl-2 dosage compensation gene of Drosophila encodes a putative
 RT DNA-binding protein whose expression is sex specifically regulated
 RT by Sex-lethal.";
 RL Development 121:3245-3258(1995).
 CC -1- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
 CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
 CC DOSAGE COMPENSATION). MSL-2 IS REQUIRED FOR TRANSLATION AND/OR
 CC STABILITY OF MSL-1 IN MALES. BOTH MSL-1 AND MSL-2 BIND TO HUNDREDS
 CC OF SITES ALONG THE MALE X CHROMOSOME, BUT NOT THE FEMALE X OR THE
 CC AUTOSOMES IN EITHER SEX.
 CC -1- SUBUNIT: MSL-2 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-1.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: THE MSL-2 PROTEIN IS ONLY PRODUCED IN MALES.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; L42553; AAA75573.1; -
 CC EMBL; Z4843; CAA8358.1; -
 CC EMBL; X89241; CAA61529.1; -
 CC FlyBase; FBgn0005616; msl-2.
 CC InterPro; IPR001841; Znf_Ring.

Query Match 6.0%; Score 96; DB 1; Length 773;
 Best Local Similarity 25.7%; Pred. No. 0.81;
 Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;
 QY 121 ERFKAIKAC---LDRNPEAFOPKMGKGLGAALKRHSKG-----CNCKRSG 163
 Db 475 KRTRLKASQAALTEPVSEVTKVQSGK-GA--LRRIRGDKKEKVKPKPKRCGIGS 531
 QY 164 -----CLKNYCEYBAKIMCSSICKIACKN-----YEES-----PER 196
 Db 532 SSNTLTTCRNSRCPCKYSKNSCAG-CHVCCKNPKHEDYVEDEDDLEDFEMPKDYPEP 590
 QY 197 KMLMSTPHYMPCGDFESSHYLSPAKFSGPPKL 228
 Db 591 MTQSEEPVVAEPROEENS--MAPPDSSAPISL 620

RESULT 5
 LMA5_MOUSE STANDARD; PRT; 3718 AA.
 AC Q61001; Q9JHQ6;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-5 chain precursor.
 GN LAMA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE OF 1-92 FROM N.A.
 RA Timpl R., Sasaki T.;
 RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
 RT chain.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 84-3718 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=96081906; PubMed=7499364;
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
 RT expression in adult mouse tissues.";
 RL J. Biol. Chem. 270:28523-28526(1995).
 RN [3]
 RP REVISIONS.
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ

IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.

- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.

- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.

- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.

- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.

- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.

- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AJ293593; CAB92955.1; -

EMBL: U37501; AAC53430.1; -

HSSP: P02468; 1TLE.

MGD: MGI:105382; Lama5.

InterPro: IPR000561; EGF-like.

InterPro: IPR001886; LamNT.

InterPro: IPR000034; Laminin_B.

InterPro: IPR002049; Laminin_EGF.

InterPro: IPR001791; Laminin_G.

Pfam: PF00052; laminin_B; 1.

Pfam: PF00053; laminin_EGF; 19.

Pfam: PF00054; laminin_G; 2.

Pfam: PF00055; laminin_Nterm; 1.

PRINTS: PR00011; EGF_LAMININ.

ProDom: PD002082; LamNT; 1.

ProDom: PD003031; Laminin_B; 1.

SMART: SM00180; EGF_Lam; 17.

SMART: SM00001; EGF_like; 2.

SMART: SM00281; LamB; 1.

SMART: SM00282; LamG; 5.

SMART: SM00136; LamNT; 1.

PROSITE: PS00022; EGF_1; 19.

PROSITE: PS01186; EGF_2; 3.

PROSITE: PS01248; LAMININ_TYPE_EGF; 19.

PROSITE: PS00025; LAM_G_DOMAIN; 5.

GLYCOPROTEIN: Basement membrane; Extracellular matrix; Coiled coil;

Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

SIGNAL 1 40 POTENTIAL.

FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.

FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).

FT DOMAIN 305 363 LAMININ EGF-LIKE 1.

FT DOMAIN 364 433 LAMININ EGF-LIKE 2.

FT DOMAIN 434 479 LAMININ EGF-LIKE 3.

FT DOMAIN 500 546 LAMININ EGF-LIKE 4.

FT DOMAIN 547 592 LAMININ EGF-LIKE 5.

FT DOMAIN 593 637 LAMININ EGF-LIKE 6.

FT DOMAIN 638 682 LAMININ EGF-LIKE 7.

FT DOMAIN 683 728 LAMININ EGF-LIKE 8.

FT DOMAIN 729 781 LAMININ EGF-LIKE 9.

FT DOMAIN 782 833 LAMININ EGF-LIKE 10.

FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).

FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).

FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.

FT DOMAIN	1489	1532	LAMININ EGF-LIKE 13.
FT DOMAIN	1533	1581	LAMININ EGF-LIKE 14.
FT DOMAIN	1582	1632	LAMININ EGF-LIKE 15.
FT DOMAIN	1633	1642	LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN	1643	1831	LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN	1832	1864	LAMININ EGF-LIKE 16 (C-TERMINAL).
FT DOMAIN	1865	1914	LAMININ EGF-LIKE 17.
FT DOMAIN	1915	1970	LAMININ EGF-LIKE 18.
FT DOMAIN	1971	2024	LAMININ EGF-LIKE 19.
FT DOMAIN	2025	2071	LAMININ EGF-LIKE 20.
FT DOMAIN	2072	2118	LAMININ EGF-LIKE 21.
FT DOMAIN	2119	2168	LAMININ EGF-LIKE 22.
FT DOMAIN	2169	2735	DOMAIN II AND I.
FT DOMAIN	2736	2933	LAMININ G-LIKE 1.
FT DOMAIN	2947	3119	LAMININ G-LIKE 2.
FT DOMAIN	3128	3296	LAMININ G-LIKE 3.
FT DOMAIN	3337	3511	LAMININ G-LIKE 4.
FT DOMAIN	3518	3689	LAMININ G-LIKE 5.
FT DOMAIN	2205	2257	COILED COIL (POTENTIAL).
FT DOMAIN	2330	2464	COILED COIL (POTENTIAL).
FT DOMAIN	2604	2621	COILED COIL (POTENTIAL).
FT DOMAIN	2639	2705	COILED COIL (POTENTIAL).
FT SITE	1723	1725	CELL ATTACHMENT SITE (POTENTIAL).
FT SITE	1839	1841	CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID	305	314	BY SIMILARITY.
FT DISULFID	307	327	BY SIMILARITY.
FT DISULFID	329	338	BY SIMILARITY.
FT DISULFID	341	361	BY SIMILARITY.
FT DISULFID	364	373	BY SIMILARITY.
FT DISULFID	366	398	BY SIMILARITY.
FT DISULFID	401	410	BY SIMILARITY.
FT DISULFID	413	431	BY SIMILARITY.
FT DISULFID	434	445	BY SIMILARITY.
FT DISULFID	436	452	BY SIMILARITY.
FT DISULFID	454	463	BY SIMILARITY.
FT DISULFID	466	476	BY SIMILARITY.
FT DISULFID	500	512	BY SIMILARITY.
FT DISULFID	502	521	BY SIMILARITY.
FT DISULFID	523	532	BY SIMILARITY.
FT DISULFID	535	544	BY SIMILARITY.
FT DISULFID	547	559	BY SIMILARITY.
FT DISULFID	549	566	BY SIMILARITY.
FT DISULFID	568	577	BY SIMILARITY.
FT DISULFID	580	590	BY SIMILARITY.
FT DISULFID	593	605	BY SIMILARITY.
FT DISULFID	595	611	BY SIMILARITY.
FT DISULFID	613	622	BY SIMILARITY.
FT DISULFID	625	635	BY SIMILARITY.
FT DISULFID	638	650	BY SIMILARITY.
FT DISULFID	640	656	BY SIMILARITY.
FT DISULFID	658	667	BY SIMILARITY.
FT DISULFID	670	680	BY SIMILARITY.
FT DISULFID	683	695	BY SIMILARITY.
FT DISULFID	685	702	BY SIMILARITY.
FT DISULFID	704	713	BY SIMILARITY.
FT DISULFID	716	726	BY SIMILARITY.
FT DISULFID	1443	1455	BY SIMILARITY.
FT DISULFID	1445	1462	BY SIMILARITY.
FT DISULFID	1464	1473	BY SIMILARITY.
FT DISULFID	1476	1486	BY SIMILARITY.
FT DISULFID	1533	1548	BY SIMILARITY.
FT DISULFID	1535	1555	BY SIMILARITY.
FT DISULFID	1557	1566	BY SIMILARITY.
FT DISULFID	1569	1579	BY SIMILARITY.
FT DISULFID	1582	1594	BY SIMILARITY.
FT DISULFID	1584	1601	BY SIMILARITY.
FT DISULFID	1603	1612	BY SIMILARITY.
FT DISULFID	1615	1630	BY SIMILARITY.
FT DISULFID	1665	1874	BY SIMILARITY.
FT DISULFID	1867	1881	BY SIMILARITY.
FT DISULFID	1884	1893	BY SIMILARITY.
FT DISULFID	1896	1912	BY SIMILARITY.
FT DISULFID	1915	1930	BY SIMILARITY.

```
FT DISULFID 1917 1939 BY SIMILARITY.
FT DISULFID 1941 1950 BY SIMILARITY.
FT DISULFID 1953 1968 BY SIMILARITY.
FT DISULFID 1971 1986 BY SIMILARITY.
FT DISULFID 1973 1993 BY SIMILARITY.
FT DISULFID 1996 2005 BY SIMILARITY.
FT DISULFID 2008 2022 BY SIMILARITY.
FT DISULFID 2072 2083 BY SIMILARITY.
FT DISULFID 2074 2090 BY SIMILARITY.
FT DISULFID 2092 2101 BY SIMILARITY.
FT DISULFID 2104 2116 BY SIMILARITY.
FT DISULFID 2119 2126 BY SIMILARITY.
FT DISULFID 2121 2133 BY SIMILARITY.
FT DISULFID 2135 2144 BY SIMILARITY.
FT DISULFID 2147 2166 BY SIMILARITY.
FT DISULFID 2169 2169 INTERCHAIN (PROBABLE).
FT DISULFID 2172 2172 INTERCHAIN (PROBABLE).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 6.0%; Score 96; DB 1; Length 3718;
Best Local Similarity 18.0%; Pred. No. 4;
Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;

QY 14 CIDNCGARELKALHLLPOYDDQSFPOSELPKPMITLVGLLPPAKNLNLTQVDNGLP 73
Db 200 CLERFGPRTLER---ITQDDVI-----CTTEYSRIVPL-ENGEIWSLVNGR-P 244
QY 74 SAVNGAAPP-----SGPALQGP-----PKITLSGY 98
Db 245 GAINFYSPLLRDTKATNIRLRLNTLTLGLHLMGKALRDPVTRRYYSIKDISIGR 304
QY 99 CDCFSSGDFC-----NSC-----SC 113
Db 305 CVCHGADVCDKPLDPFRLOCAQHNTCGGSDRCPCGPNQOPKWPATTDSANECQSC 364
QY 114 NNLRH-----ELEFEKAKACLDNRNPAFPQKMGKGLGAAL---RHSGKCNKRS 162
Db 365 NCHGHAYDCYDPEVDR-----RNASONQDNVYQG--GGVCLDCQHHTTGINGER- 412
QY 163 GCLKNY-----CEYEAKI---MCSII---CK-----CIAC-KNY 190
Db 413 -CLPFFRAPDPLDSPHVCRPCDC-ESDFTDGTCDLTGRCYCRPNFTGELCAACAGY 470
QY 191 EESPERKMLMSTHYMEPGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWEVVEATCACL 250
Db 471 TDFPHCYPLPSFPHN-----DTREQVLPA-----GQIVNCDCAA 505
QY 251 LAQEEAEQE-----HC-----SPLAEQMLEFEGR 277
Db 506 GTQGNACKRDPRLGRGCKVKNFNGAHGELCAPGPHGSPCHPQCQSSPGVANSLCDPESGQ 565
QY 278 CL 279
Db 566 CM 567
```

RESULT 6

```
ID KPC1_ASPNG STANDARD; PRT; 1096 AA.
AC Q00078;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein kinase C-like (EC 2.7.1.-).
GN PKCA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
```

```
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
RN NCBI_TaxID=5061;
RP SEQUENCE FROM N.A.
RX MEDLINE=96158841; PubMed=8569684;
RA Morawetz R., Lendenfeld T., Mischak H., Muehlbauer M., Gruber F.,
RA Goodnight J., de Graaff L.H., Visser J., Mushinski J.F.,
RA Kubicek C.P.;
RT "Cloning and characterisation of genes (pkc1 and pkcA) encoding
RT protein kinase C homologues from Trichoderma reesei and Aspergillus
RT niger.";
RL Mol. Gen. Genet. 250:17-28(1996).
CC -|- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -|- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC PKC SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U10549; AAA37433.1; -
CC HSP; Q63450; 1A06.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000961; Pkinase_C.
DR InterPro; IPR000861; REM_repeat.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00130; DAG_PE-blind; 2.
DR Pfam; PF02185; HRI; 2.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00433; pkinase_C; 1.
DR SMART; SM00109; C1; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00074; HRI; 2.
DR SMART; SM00133; S_TK_X; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; zinc;
KW Phorbol-ester binding; Repeat.
FT DOMAIN 460 507 PHORBOL-ESTER AND DAG BINDING 1.
FT DOMAIN 528 577 PHORBOL-ESTER AND DAG BINDING 2.
FT DOMAIN 771 1030 PROTEIN KINASE.
FT NP_BIND 777 785 ATP (BY SIMILARITY).
FT BINDING 800 800 ATP (BY SIMILARITY).
FT ACT_SITE 896 896 BY SIMILARITY.
SQ SEQUENCE 1096 AA; 122234 MW; 859B2DB3B3D5B08 CRC64;
```

Query Match 5.9%; Score 95; DB 1; Length 1096;

Best Local Similarity 22.6%; Pred. No. 1.5;

Matches 63; Conservative 21; Mismatches 91; Indels 104; Gaps 15;

```
QY 66 QVDNGALPSAVN-----GAAPSPGALQGPPIKITLSGYCDGFCSSGDFCSCNNRLHEL 120
Db 359 KMEHGAGRPDTAGGAGSAHPSGP--GPAAPSGYDGSAGAPAGNQVMDISW-FAL 415
QY 121 ERFKAI-----KACLDNRNPEAFQPKMGKGLGAALKR-----HSKG-----C 157
Db 416 EPVGRYLSMSFAKQLKDRRPF---DIGLNQCAVVRQKKEVEHKGQHKFVTQQFYNIM 471
QY 158 NKRSGCLKNYC---ECYEAKIMCSSIC-----KCIACKNVEESPERKMLMSTPHYMEP 208
Db 472 RCALCGDFLKYAAGMCQACDKCYTHKKCYPKVVTYTKISKANYETDP----- 517
```


FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1086 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1281 1281 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1359 1359 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1411 1411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 143 143 A -> P (IN REF. 1).
 FT CONFLICT 178 178 L -> F (IN REF. 1).
 FT CONFLICT 265 265 G -> GMDPTIS (IN REF. 4).
 FT CONFLICT 276 276 D -> A (IN REF. 4).
 FT CONFLICT 491 491 Y -> H (IN REF. 2 AND 3).
 FT CONFLICT 1057 1057 T -> P (IN REF. 1).
 FT CONFLICT 1110 1112 SGR -> GGP (IN REF. 4).
 SQ SEQUENCE 1816 AA; 201908 MW; 04E9AF379A0F4A4D CRC64;

Query Match 5.8%; Score 93; DB 1; Length 1816;

Best Local Similarity 23.8%; Pred. No. 3.8;
 Matches 74; Conservative 32; Mismatches 143; Indels 62; Gaps 18;

QY 2 VICQLKGQMLCTDNGC-----ARELKALH---LLPOY--DDQSGFPQSELKPKMTTLV 51
 DB 80 VPDCNGNSNE-CLDGGYCVHCORNTTGEHCEKCLDGYGDSIRGAPQCPCP----- 133
 QY 52 GRLLVPKALNLTQV--DNGALPSAVN-GAAPPSS-----GPALQGGPKITLSGYCDCFSS 104
 DB 134 --CPLPLHNAFAESCYRKNAGVRCICNENYAGPNCERCAPGYGNPLLI----- 180
 QY 105 GDFCNSCSN-----NLRHE--LERFAIKACLDNRPEAQ-PYMGKGRGLGAAL-RHSK 155
 DB 181 GSTCKKDCSGNDPNLIFEDCDVTGQCRNCL-RNTTGKERCAPGYGYGDIARIAKCA 239
 QY 156 GCNCKRSGCLKNYCEVCEAKMSSIC-KCI-----ACKNYEESPERKMLMST--P 203
 DB 240 VCMGGGFCDSVTCGELEEGFEPTGCKVCWDLTDRLAALSIEGKSGVLSVSGAA 299
 QY 204 HYMEPGDFESSHSPAKFSGPKLRNQAFCISWEVVEATCCLLAQGEAEQEHCS 263
 DB 300 AHRHVEINATVLLKTLSE----RENOVALRKIQINNAENTMKSLSVDVELKEQENQ 355
 QY 264 PSIAEQMILEE 274
 DB 356 ASRKQLVQKE 366

RESULT 8

ID ITBL_XENLA STANDARD; PRT; 798 AA.
 AC P12606;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Integrin beta-1 precursor.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88186829; PubMed=28333505;
 RA Desimone D.W., Hynes R.O.;
 RT "Xenopus laevis integrins. Structural conservation and evolutionary
 divergence of integrin beta subunits.";
 RL J. Biol. Chem. 263:5333-5340(1988).
 CC -1- FUNCTION: BETA INTEGRINS ASSOCIATE WITH ALPHA SUBUNITS TO FORM
 RECEPTOR COMPLEXES THAT RECOGNIZE THE SEQUENCE R-G-D IN A WIDE
 CC ARRAY OF LIGANDS (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND BETA SUBUNIT.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY. THE
 INTEGRINS BETA-1 AND BETA-1* OF XENOPUS ONLY DIFFER BY 19 AMINO

CC ACIDS.
 CC -1- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; M20140; AAA49889.1; -.
 CC PIR; A28193; A28193.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002369; Integrin_B.
 DR InterPro; IPR001169; Integrin_beta_C.
 DR InterPro; IPR003659; PSI.
 DR InterPro; IPR002035; VWFA.
 DR Pfam; PF00362; Integrin_B; 1.
 DR PRINTS; PR01186; INTEGRINB.
 DR ProDom; PD001811; Integrin_B; 1.
 DR SMART; SM00187; INB; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00243; INTEGRIN_BETA; 3.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
 KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
 KW Repeat; Signal; Phosphorylation.
 FT SIGNAL 1 21
 FT CHAIN 22 798 INTEGRIN BETA-1.
 FT DOMAIN 22 727 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 728 751 POTENTIAL.
 FT DOMAIN 752 798 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 139 377 VWFA-LIKE.
 FT DOMAIN 466 635 4 CYSTEINE-RICH TANDEM REPEATS.
 FT REPEAT 466 535 I.
 FT REPEAT 516 559 II.
 FT REPEAT 560 598 III.
 FT REPEAT 599 635 IV.
 FT DISULFID 28 464 BY SIMILARITY.
 FT DISULFID 36 46 BY SIMILARITY.
 FT DISULFID 39 76 BY SIMILARITY.
 FT DISULFID 49 65 BY SIMILARITY.
 FT DISULFID 206 212 BY SIMILARITY.
 FT DISULFID 260 300 BY SIMILARITY.
 FT DISULFID 400 414 BY SIMILARITY.
 FT DISULFID 434 691 BY SIMILARITY.
 FT DISULFID 462 466 BY SIMILARITY.
 FT DISULFID 477 489 BY SIMILARITY.
 FT DISULFID 486 525 BY SIMILARITY.
 FT DISULFID 491 500 BY SIMILARITY.
 FT DISULFID 502 516 BY SIMILARITY.
 FT DISULFID 531 536 BY SIMILARITY.
 FT DISULFID 533 568 BY SIMILARITY.
 FT DISULFID 538 553 BY SIMILARITY.
 FT DISULFID 555 560 BY SIMILARITY.
 FT DISULFID 574 579 BY SIMILARITY.
 FT DISULFID 576 607 BY SIMILARITY.
 FT DISULFID 581 590 BY SIMILARITY.
 FT DISULFID 592 599 BY SIMILARITY.
 FT DISULFID 613 618 BY SIMILARITY.
 FT DISULFID 615 661 BY SIMILARITY.
 FT DISULFID 620 630 BY SIMILARITY.
 FT DISULFID 633 636 BY SIMILARITY.
 FT DISULFID 640 649 BY SIMILARITY.
 FT DISULFID 646 723 BY SIMILARITY.
 FT DISULFID 665 699 BY SIMILARITY.
 FT MOD_RES 783 783 PHOSPHORYLATION (BY SIMILARITY).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 798 AA; 88167 MW; 093522509B298531 CRC64;

Query Match
Best Local Similarity 5.8%; Score 92; DB 1; Length 798;
Matches 47; Conservative 14; Mismatches 71; Indels 82; Gaps 12;

QY 95 LSGYCD-----CFSSGD-FCNCSNNLRHELFKRAIKACLDNRNPEAFQPMKGRL 146
DQ 512 MDAYCRRENSSEICSNCGICGQCCKRONPNVEYSG-KYCECDN---FNCDSRNLGI 567
QY 147 GAAKLHRSKG-CNCKRSGLKNY-----CEYEAKIMCSS-----ICKIACKNVE 191
DQ 568 CG-----GKGICKRVCECFNPGSGSACDCSDTSTCMKNGQICNGRIGCDGRCK--- 619
QY 192 ESPERKMLMSTPHYMPEGDFESSHVSFPAKSGPPKLRKNRQAFSCISWEVVEATC-ACL 250
DQ 620 -----CTDPKFGP-----TCLELQ 634

QY 251 LAQGEAEQHCSPSLAQMIIEFGRLSQTILH 284
DQ 635 TCVGVCAEHKEVCQCFQK-GEQDVCMQCMH 667

RESULT 9
OC90_HUMAN
AC 002509; STANDARD; PRT; 493 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 01-MAR-2002 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Otc90in 90 precursor (Otc90) (Phospholipase A2 homolog).
GN OTC90 OR PLA2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93181166; PubMed=8382789;
RA Feuchter-Nurthy A.E., Freeman J.D., Mager D.L.;
RT "Splicing of a human endogenous retrovirus to a novel phospholipase
RL Nucleic Acids Res. 21:135-143(1993).
CC -!- FUNCTION: IT IS UNLIKELY THAT THIS PROTEIN HAS PHOSPHOLIPASE A2
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. CONTAINS 3
CC PA2 TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z14310; CAA78662.1; ALT_INIT.
CC HSP; P00593; 4BP2.
CC MIN; 601658;
CC InterPro; IPR001211; PLP_A2.
CC Pfam; PF00068; phoslip; 3.
CC PRINTS; PR00389; PHPLIPASEA2.
CC ProDom; PD000303; PLP_A2; 2.
CC SMART; SM00085; PA2C; 2.
CC PROSITE; PS00118; PA2_HIS; 2.
```

```
DR PROSITE; PS00119; PA2_ASP; 1.
KW Signal; Glycoprotein; Repeat.
FT SIGNAL 1 17 POTENTIAL..
FT CHAIN 18 493 OTOCONIN 90.
FT DOMAIN 76 190 PHOSPHOLIPASE A2-LIKE 1.
FT DOMAIN 321 377 PHOSPHOLIPASE A2-LIKE 2.
FT DOMAIN 389 441 PHOSPHOLIPASE A2-LIKE 3.
FT DISULFID 85 145 BY SIMILARITY.
FT DISULFID 99 190 BY SIMILARITY.
FT DISULFID 101 117 BY SIMILARITY.
FT DISULFID 116 172 BY SIMILARITY.
FT DISULFID 123 165 BY SIMILARITY.
FT DISULFID 132 158 BY SIMILARITY.
FT DISULFID 152 163 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 493 AA; 53338 MW; 59F9AF7C1364A5B7 CRC64;

Query Match
Best Local Similarity 5.7%; Score 91; DB 1; Length 493;
Matches 59; Conservative 50; Mismatches 89; Indels 124; Gaps 14;

QY 2 VICQLKGGAQ-MLCINDGARELKA-----LHLLPOYDDQSPFQSELKPMTTLVGR 53
DQ 150 IICESKDNCEHLCTCDKAAIECLARSSLSNLLDTSFCLAQTPETTIKEDLTLLPR 209
QY 54 LLPV-PAKLNI-----TQVD-----NGALPSAVNGAA----- 80
DQ 210 VVPVEPTDTSLTALSGEVAATEADRLITLSKKAGHDQEGVGAARATSPPGSAEIVATR 269
QY 81 -----FPSG-----PALQPPKTLTSGYDCDFSSGDFCNSCNSNLRHELERK 124
DQ 270 VTAKITLVPAGIKSLGLAVSSVENDPEETTEKACDRFT---FLHLSGDNHQMVPOLGE 326
QY 125 AIAKCLDRNPEAFQPMKGRLGAKLRHSGKCNKRSQ-----CLKNYCECYE 173
DQ 327 MLFCLTSRCPEFE-----SYCYCGQEGEGRDRDLDRCLLSHHC-CLE 370
QY 174 -----AKIMCS-----SICKIACKNVEEESPERKM 198
DQ 371 QVRRGLGCLLERLPWSPVVDHTPRKGGGSLCEKLLCACDQTAACMTSASFNSQLKSPS 430
QY 199 LMSTPHYMPEGDFESSHYLSPA 220
DQ 431 RIGCGP--OPACEDSLHPVPA 450

RESULT 10
EZHL_HUMAN
ID EZHL_HUMAN STANDARD; PRT; 747 AA.
AC Q92800; O43287; Q14459;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enhancer of zeste homolog 1 (ENX-2).
GN EZHL OR KIAA0388.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97079676; PubMed=8921387;
RA Abel K.J., Brody L.C., Valdes J.M., Erdos M.R., McKinley D.R.,
RA Castella L.H., Merajver S.D., Couch F.J., Friedman L.S.,
RA Ostermeyer E.A., Lynch E.D., King M.-C., Welch P.L.,
RA Osborne-Lawrence S., Spillman M., Bowcock A.M., Collins F.S.,
RA Weber B.L.;
RT "Characterization of EZHL, a human homolog of Drosophila Enhancer of
RT zeste near BRCA1."
RL Genomics 37:161-171(1996).
```

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98146265; PubMed=9473645;
 RA Ogawa M., Hiroaka Y., Taniguchi K., Aiso S.;
 RT "Cloning and expression of a human/mouse Polycomb group gene,
 RL ENX-2/Enx-2.";
 RL Biochim. Biophys. Acta 1395:151-158(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [4]
 RP SEQUENCE OF 434-538 FROM N.A.
 RX MEDLINE=96039267; PubMed=7490091;
 RA Rommens J.M., Durocher F., McArthur J., Tonin P., Leblanc J.F.,
 RA Allen T., Samson C., Ferri L., Narod S., Morgan K., Simard J.;
 RT "Generation of a transcription map at the HSD17B locus centromeric to
 RT BRCA1 at 17q21.";
 RL Genomics 28:530-542(1995).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF GENE TRANSCRIPTION
 CC AND CHROMATIN STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE EZ FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U50315; AAC50778.1;
 CC EMBL: AB004818; BAA25019.1;
 CC EMBL: AB002386; BAA20842.1;
 CC EMBL: L38934; AAB59574.1;
 CC MIM: 601674;
 CC InterPro: IPR001719; AP_endonuclease_2.
 CC InterPro: IPR001005; Myb_DNA_bind.
 CC InterPro: IPR001214; SET.
 CC Pfam: PF00836; SET; 1.
 CC SMART: SM00395; SANT; 1.
 CC SMART: SM00317; SET; 1.
 CC PROSITE: PS50280; SET; 1.
 CC Transcription regulation; Nuclear protein; DNA-binding.
 FT DOMAIN 491 496 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 524 606 CYS-RICH.
 FT DOMAIN 619 730 SET.
 FT CONFLICT 353 353 P -> S (IN REF. 2).
 FT CONFLICT 488 488 N -> V (IN REF. 1).
 FT CONFLICT 532 535 DSTC -> EAL (IN REF. 2).
 FT CONFLICT 591 602 ASHWCKVVC -> POSTGTARWPPV (IN REF. 2).
 FT CONFLICT 700 747 VVMVGDHIGIFAKRAIQAGEELFFDYRSQADALKYGI
 FT ERETDLV -> GESQ (IN REF. 2).
 SQ SEQUENCE 747 AA; 85270 MW; 7CFC52269CDA011B CRC64;

 Query Match 5.7%; Score 91; DB 1; Length 747;
 Best local Similarity 21.0%; Pred. No. 2.1;
 Matches 37; Conservative 17; Mismatches 44; Indels 78; Gaps 9;

 QY 99 CDCFSSGDFCNS-CSCNNLRHELFRFAIKRAIDLRNPEAFQPMKGKRLGAALKRHSKGC 157
 Db 535 CPCIMTQNFCEKFCQC-----PCQNRFP-----GC 561
 QY 158 NCRSGCLKNYCEYAKIMCSS-----ICKIACKNYE-ESPERKMLMSTPH 204

Db 562 RCK-TQCNKQPCYLAVERCDPDLCLTCGASEHWDCKVVSCKNCSIQRLKHL----- 616
 QY 205 YMEPGDFESSHYLSPAKFSG-----PPKLRKNRQAFSCISWEVATCACLLAQGE 255
 Db 617 -----LAPSDVAGWGTFFIKESQKN-----EFISEYCGELISQDE 651

 RESULT 11;
 FBN2_MOUSE
 ID FBN2_MOUSE STANDARD; PRT; 2907 AA.
 AC Q61555; Q63957;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Fibrillin 2 precursor.
 GN FBN2 OR FBN-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95263670; PubMed=7744963;
 RA Zhang H., Hu W., Ramirez F.;
 RT "Developmental expression of fibrillin genes suggests heterogeneity
 RT of extracellular microfibrils.";
 RL J. Cell Biol. 129:1165-1176(1995).
 RN [2]
 RP SEQUENCE OF 210-317 FROM N.A.
 RX MEDLINE=94140368; PubMed=8307578;
 RA Li X., Pereira L., Zhang H., Sanguineti C., Ramirez F., Bonadio J.,
 RA Francke U.;
 RT "Fibrillin genes map to regions of conserved mouse/human synteny on
 RT mouse chromosomes 2 and 18.";
 RL Genomics 18:667-672(1993).
 CC -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
 CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
 CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
 CC -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
 CC EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L39790; AAA74908.1;
 CC EMBL: S69359; AAC60685.1;
 CC HSP: P35555; 1EMN
 CC MGD; MGI:95490; Fbn2.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR001438; EGF_II.
 CC InterPro: IPR002212; TB.
 CC InterPro: IPR000822; Znf-C2H2.
 CC Pfam: PF00008; EGF; 46.
 CC Pfam: PF00683; TB; 9.
 CC PRINTS; PR00010; EGFBL00D.
 CC SMART; SM00179; EGF_CA; 43.
 CC SMART; SM00001; EGF_Like; 3.
 CC PROSITE; PS00010; ASX_HYDROXYL; 43.
 CC PROSITE; PS00022; EGF_1; 2.
 CC PROSITE; PS01186; EGF_2; 36.
 CC PROSITE; PS01187; EGF_CA; 43.
 CC Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
 KW Repeat; Signal; Multigene family.
 FT SIGNAL 1 28 POTENTIAL.

FT	DISULFID	532	542	BY SIMILARITY.
FT	DISULFID	537	551	BY SIMILARITY.
FT	DISULFID	553	566	BY SIMILARITY.
FT	DISULFID	572	584	BY SIMILARITY.
FT	DISULFID	579	593	BY SIMILARITY.
FT	DISULFID	595	608	BY SIMILARITY.
FT	DISULFID	614	625	BY SIMILARITY.
FT	DISULFID	620	634	BY SIMILARITY.
FT	DISULFID	636	649	BY SIMILARITY.
FT	DISULFID	655	666	BY SIMILARITY.
FT	DISULFID	661	675	BY SIMILARITY.
FT	DISULFID	677	690	BY SIMILARITY.
FT	DISULFID	765	777	BY SIMILARITY.
FT	DISULFID	772	786	BY SIMILARITY.
FT	DISULFID	788	801	BY SIMILARITY.
FT	DISULFID	807	819	BY SIMILARITY.
FT	DISULFID	814	828	BY SIMILARITY.
FT	DISULFID	830	843	BY SIMILARITY.
FT	DISULFID	849	859	BY SIMILARITY.
FT	DISULFID	854	868	BY SIMILARITY.
FT	DISULFID	870	883	BY SIMILARITY.
FT	DISULFID	952	964	BY SIMILARITY.
FT	DISULFID	959	973	BY SIMILARITY.
FT	DISULFID	975	988	BY SIMILARITY.
FT	DISULFID	1070	1082	BY SIMILARITY.
FT	DISULFID	1077	1091	BY SIMILARITY.
FT	DISULFID	1093	1106	BY SIMILARITY.
FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1119	1133	BY SIMILARITY.
FT	DISULFID	1135	1149	BY SIMILARITY.
FT	DISULFID	1155	1167	BY SIMILARITY.
FT	DISULFID	1162	1176	BY SIMILARITY.
FT	DISULFID	1178	1191	BY SIMILARITY.
FT	DISULFID	1197	1209	BY SIMILARITY.
FT	DISULFID	1204	1218	BY SIMILARITY.
FT	DISULFID	1220	1233	BY SIMILARITY.
FT	DISULFID	1239	1250	BY SIMILARITY.
FT	DISULFID	1245	1259	BY SIMILARITY.
FT	DISULFID	1261	1274	BY SIMILARITY.
FT	DISULFID	1280	1292	BY SIMILARITY.
FT	DISULFID	1287	1301	BY SIMILARITY.
FT	DISULFID	1303	1316	BY SIMILARITY.
FT	DISULFID	1322	1334	BY SIMILARITY.
FT	DISULFID	1329	1343	BY SIMILARITY.
FT	DISULFID	1345	1358	BY SIMILARITY.
FT	DISULFID	1364	1377	BY SIMILARITY.
FT	DISULFID	1371	1386	BY SIMILARITY.
FT	DISULFID	1388	1399	BY SIMILARITY.
FT	DISULFID	1405	1418	BY SIMILARITY.
FT	DISULFID	1412	1427	BY SIMILARITY.
FT	DISULFID	1429	1440	BY SIMILARITY.
FT	DISULFID	1446	1458	BY SIMILARITY.
FT	DISULFID	1453	1467	BY SIMILARITY.
FT	DISULFID	1469	1482	BY SIMILARITY.
FT	DISULFID	1488	1499	BY SIMILARITY.
FT	DISULFID	1494	1508	BY SIMILARITY.
FT	DISULFID	1510	1523	BY SIMILARITY.
FT	DISULFID	1529	1540	BY SIMILARITY.
FT	DISULFID	1535	1549	BY SIMILARITY.
FT	DISULFID	1551	1564	BY SIMILARITY.
FT	DISULFID	1647	1659	BY SIMILARITY.
FT	DISULFID	1654	1668	BY SIMILARITY.
FT	DISULFID	1670	1683	BY SIMILARITY.
FT	DISULFID	1689	1701	BY SIMILARITY.

Db 405 CPVRGSEYYRRLCLDG-----LP-----MGGIPGSSSVSRPGGT--GSNGNGYGP 447

QY 54 ----LLPVPKALNLTQVDNGALPSAVNGAARFSGPALQGP-----PK 92
D 448 GTGFLPIPG-----DNGFSP-CVGAGVAGG--GGPIITGLTILNQTDICKHHAN 496
QY 93 ITLSGYC-----DCP-----SSGDFCNS-----CSCNNL 116
D 497 LCLNGRCITVSSYRCRCNMGYKQDANGCDIDVDECTSNPCSNQGVNTPGSYCKCH-- 554
QY 117 RHELEFRAIKACLDNRPEAFQPKMGKRLGAALKHSGKNCRGKGLK----- 166
D 555 -AGFORTPTKQACIDID-ECIQ-----NGVLCKNGRCVNSDGSFQCICN 596
QY 167 -----NYCEYEAKIMC-----KCIACKNYBESPERKMLMSTPHYM 206
D 597 AGFELTTDGNKNCVDHDECTTNNCLNGMCINEDSGFKVCVCPGFILAPNGRYCTDWDCEQ 656
QY 207 EPGDEFSSHYSL---PAKFSPPKLRKRRQAFSCIS-----WEVVEATCA----- 248
D 657 TPGICMNGHCINNEGSFRCDPCPPGLAVGVDRVCVDTMRSTCYGEIKRGVCVRPFGAV 716
QY 249 ----CLLAGGEAEQEHCSPLAEQMILEEGRCLSQI 282
D 717 TKSECCANPDYGFGEPCPCPAKNS-AEFHGLCSSGI 753
RESULT 12
LMB1_MOUSE
AC P02469; STANDARD; PRT; 1786 AA.
DT 01-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LAMB1-1 OR LAMB-1
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87147212; PubMed=3493487;
RA Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;
RT "Sequence of the cDNA encoding the laminin B1 chain reveals a
multidomain protein containing cysteine-rich repeats.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).
RN [2]
RP SEQUENCE OF 1292-1786 FROM N.A.
RX MEDLINE=85051302; PubMed=6209134;
RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
coiled-coil alpha-helix.";
RL EMBO J. 3:2355-2362(1984).
RN [3]
RP SEQUENCE OF 165-172; 539-547 AND 712-719.
RX STRAIN=BAUB/C; TISSUE=Endothelial cells;
RC MEDLINE=97363207; PubMed=9219532;
RA Frieser M., Noeckel H., Pausch C., Roeder C., Hahn A., Deutzmann R.,
RA Sorokin L.M.;
RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of
endothelium.";
RL Eur. J. Biochem. 246:727-735(1997).
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-
2 (MEROSIN), AND LAMININ-6 (K-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR

CC COMPONENT).
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M15525; AAA39407.1; ALT_INIT.
CC EMBL; X05212; CAA28839.1; -
CC PIR; A26413; MMSB1.
CC HSP; P02468; IKLO.
CC MGD; MGI:96743; Lamb1-1.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR002049; Laminin_EGF.
CC Pfam; PF00053; laminin_EGF; 13.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRODOM; PD002082; LamNT; 1.
CC SMART; SM00180; EGF_Lam; 11.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 9.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC
CC SIGNAL 1 21 LAMININ BETA-1 CHAIN
CC CHAIN 22 1786 LAMININ N-TERMINAL (DOMAIN VI).
CC DOMAIN 271 334 LAMININ EGF-LIKE 1.
CC DOMAIN 335 397 LAMININ EGF-LIKE 2.
CC DOMAIN 398 457 LAMININ EGF-LIKE 3.
CC DOMAIN 458 509 LAMININ EGF-LIKE 4.
CC DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).
CC DOMAIN 541 772 LAMININ DOMAIN IV.
CC DOMAIN 773 820 LAMININ EGF-LIKE 6.
CC DOMAIN 821 866 LAMININ EGF-LIKE 7.
CC DOMAIN 867 916 LAMININ EGF-LIKE 8.
CC DOMAIN 917 975 LAMININ EGF-LIKE 9.
CC DOMAIN 976 1027 LAMININ EGF-LIKE 10.
CC DOMAIN 1028 1083 LAMININ EGF-LIKE 11.
CC DOMAIN 1084 1131 LAMININ EGF-LIKE 12.
CC DOMAIN 1132 1178 LAMININ EGF-LIKE 13.
CC DOMAIN 1179 1397 DOMAIN II.
CC DOMAIN 1398 1430 DOMAIN ALPHA.
CC DOMAIN 1431 1786 DOMAIN I.
CC DOMAIN 1787 1835 COILED COIL (POTENTIAL).
CC DOMAIN 1836 1888 COILED COIL (POTENTIAL).
CC DISULFID 271 280 BY SIMILARITY.
CC DISULFID 273 298 BY SIMILARITY.
CC DISULFID 300 309 BY SIMILARITY.
CC DISULFID 312 332 BY SIMILARITY.
CC DISULFID 335 344 BY SIMILARITY.
CC DISULFID 337 362 BY SIMILARITY.
CC DISULFID 365 374 BY SIMILARITY.
CC DISULFID 377 395 BY SIMILARITY.
CC DISULFID 398 411 BY SIMILARITY.
CC DISULFID 400 426 BY SIMILARITY.
CC DISULFID 428 437 BY SIMILARITY.
CC DISULFID 440 455 BY SIMILARITY.
CC DISULFID 458 472 BY SIMILARITY.
CC DISULFID 460 479 BY SIMILARITY.
CC DISULFID 481 490 BY SIMILARITY.
CC DISULFID 493 507 BY SIMILARITY.
CC DISULFID 773 785 BY SIMILARITY.
CC DISULFID 775 792 BY SIMILARITY.


```
FT DISULFID 794 803 BY SIMILARITY.
FT DISULFID 806 818 BY SIMILARITY.
FT DISULFID 821 833 BY SIMILARITY.
FT DISULFID 823 840 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 864 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 869 883 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 898 914 BY SIMILARITY.
FT DISULFID 917 933 BY SIMILARITY.
FT DISULFID 919 944 BY SIMILARITY.
FT DISULFID 946 955 BY SIMILARITY.
FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 978 997 BY SIMILARITY.
FT DISULFID 1000 1009 BY SIMILARITY.
FT DISULFID 1012 1025 BY SIMILARITY.
FT DISULFID 1084 1096 BY SIMILARITY.
FT DISULFID 1105 1114 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1132 1144 BY SIMILARITY.
FT DISULFID 1134 1151 BY SIMILARITY.
FT DISULFID 1153 1162 BY SIMILARITY.
FT DISULFID 1165 1176 BY SIMILARITY.
FT DISULFID 1179 1179 INTERCHAIN (PROBABLE).
FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).
FT DISULFID 1185 1185 INTERCHAIN (PROBABLE).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1041 1041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1195 1195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1279 1279 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1343 1343 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1533 1533 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1542 1542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1643 1643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1531 1534 SGNA -> MEMP (IN REF. 2).
FT CONFLICT 1749 1749 D -> N (IN REF. 2).
SQ SEQUENCE 1786 AA; 196904 MW; 846671B7BF41A474 CRC64;

Query Match 5.6%; Score 89; DB 1; Length 1786;
Best Local Similarity 19.5%; Pred. NO. 8.3;
Matches 50; Conservative 32; Mismatches 102; Indels 72; Gaps 12;

QY 82 PSGP-----ALQGPKITLS-----GY----CDCFSSGDF-----CHSCSCN 114
DB 920 PDGDSGQGFARSCYQDPVTTLQACVDPGYIGSRCDGACSGFFGNPSDFGSGCQPCQCH 979
QY 115 NLRLHELRFKAKACLDNRPEAFQPKMGR-----GRLGAALRHSGK 156
DB 980 ----HNID-----TTDPEACDRDTGRCLAKLYHTGDHCLQCYGYGDLAQDCRK 1027
QY 157 CNCKRSGLKNY-----CECYEAKTMCISICKIACKNVEESPERKMLMSTPHYEPGDF 211
DB 1028 CVCNLTGTVKEHCNSDCHDKATQCSCLPNVIG-QNCDRCAPNTWQLASGTGCGPCNC 1086
QY 212 ESSHYLSA--KFSPPKLRKKRQAFSC-----ISEWVEATCCLLAQGEAEQEHSCSP 264
DB 1087 NAAHSFGPSCNFTGQCQCPMGFGGRTGTCSEQLFWGDDVECRACDCDPRGIETPQCQDQ 1146
QY 265 SLAEQMILE--EFGRC 278
DB 1147 STGQCVCVGEVGEPRC 1162

RESULT 13
```

```
ITB0_XENLA
ID ITB0_XENLA STANDARD; PRT; 798 AA.
AC P12607;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin beta-1* precursor.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88186829; PubMed=28333505;
RA Desimone D.W., Hynes R.O.; Structural conservation and evolutionary
RT divergence of integrin beta subunits.*;
RL J. Biol. Chem. 263:5333-5340(1988).
CC -!- FUNCTION: BETA INTEGRINS ASSOCIATE WITH ALPHA SUBUNITS TO FORM
CC RECEPTOR COMPLEXES THAT RECOGNIZE THE SEQUENCE R-G-D IN A WIDE
CC ARRAY OF LIGANDS (BY SIMILARITY).
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND BETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY. THE
CC INTEGRINS BETA-1 AND BETA-1* OF XENOPUS ONLY DIFFER BY 19 AMINO
CC ACIDS.
CC -!- SIMILARITY: CONTAINS 1 WFA-LIKE DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20180; AAA49890.1; .
CC PIR; B28193; B28193
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002369; Integrin_B.
CC InterPro; IPR001169; Integrin_beta_C.
CC InterPro; IPR003659; PSI.
CC InterPro; IPR002035; WFA.
CC Pfam; PF00362; Integrin_B; 1.
CC PRINTS; PR01186; INTEGRINB.
CC ProDom; PD001811; Integrin_B; 1.
CC SMART; SM00187; INB; 1.
CC SMART; SM00423; PSI; 1.
CC SMART; SM00327; WFA; 1.
CC PROSITE; PS00243; INTEGRIN_BETA; 3.
CC PROSITE; PS00222; EGF_1; UNKNOWN_2.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Signal; Phosphorylation.
FT SIGNAL 1 21
FT CHAIN 22 798 INTEGRIN BETA-1*.
FT DOMAIN 22 727 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 728 751 POTENTIAL.
FT DOMAIN 752 798 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 139 377 WFA-LIKE.
FT DOMAIN 466 635 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 466 515 I.
FT REPEAT 516 559 II.
FT REPEAT 560 598 III.
FT REPEAT 599 635 IV.
FT DISULFID 28 464 BY SIMILARITY.
FT DISULFID 36 46 BY SIMILARITY.
FT DISULFID 39 76 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 206 212 BY SIMILARITY.
FT DISULFID 260 300 BY SIMILARITY.
FT DISULFID 400 414 BY SIMILARITY.
FT DISULFID 434 691 BY SIMILARITY.
```


DR EMBL; M61945; AAA59486.1; JOINED.
DR EMBL; M61946; AAA59486.1; JOINED.
DR EMBL; M61947; AAA59486.1; JOINED.
DR EMBL; M61948; AAA59486.1; JOINED.
DR EMBL; M61949; AAA59486.1; JOINED.
DR EMBL; M61950; AAA59486.1; JOINED.
DR EMBL; M53370; AAA59485.1; JOINED.
DR EMBL; M53378; AAA59485.1; JOINED.
DR EMBL; M53371; AAA59485.1; JOINED.
DR EMBL; M53372; AAA59485.1; JOINED.
DR EMBL; M53373; AAA59485.1; JOINED.
DR EMBL; M53374; AAA59485.1; JOINED.
DR EMBL; M53375; AAA59485.1; JOINED.
DR EMBL; M53376; AAA59485.1; JOINED.
DR EMBL; M53344; AAA59485.1; JOINED.
DR EMBL; M53345; AAA59485.1; JOINED.
DR EMBL; M53346; AAA59485.1; JOINED.
DR EMBL; M53347; AAA59485.1; JOINED.
DR EMBL; M53348; AAA59485.1; JOINED.
DR EMBL; M53349; AAA59485.1; JOINED.
DR EMBL; M53350; AAA59485.1; JOINED.
DR EMBL; M53351; AAA59485.1; JOINED.
DR EMBL; M53352; AAA59485.1; JOINED.
DR EMBL; M53353; AAA59485.1; JOINED.
DR EMBL; M53355; AAA59485.1; JOINED.
DR EMBL; M53356; AAA59485.1; JOINED.
DR EMBL; M53357; AAA59485.1; JOINED.
DR EMBL; M53358; AAA59485.1; JOINED.
DR EMBL; M53359; AAA59485.1; JOINED.
DR EMBL; M53360; AAA59485.1; JOINED.
DR EMBL; M53361; AAA59485.1; JOINED.
DR EMBL; M53362; AAA59485.1; JOINED.
DR EMBL; M53363; AAA59485.1; JOINED.
DR EMBL; M53364; AAA59485.1; JOINED.
DR EMBL; M53366; AAA59485.1; JOINED.
DR EMBL; M53367; AAA59485.1; JOINED.
DR EMBL; M53368; AAA59485.1; JOINED.
DR EMBL; M53369; AAA59485.1; JOINED.
DR EMBL; M61916; AAA59482.1; -
DR EMBL; M20206; AAA59487.1; -
DR PIR; S13347; MMHUB1.
DR HSP; P02468; IKLO.
DR MIN; 150240; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00053; laminin_EGF; 13.
DR PRINTS; PR00011; EGF_LAMININ.
DR ProDom; PD002082; LamNT; 1.
DR SMART; SM00180; EGF_Lam; 11.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 9.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.
FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 271 334 LAMININ EGF-LIKE 1.
FT DOMAIN 335 397 LAMININ EGF-LIKE 2.
FT DOMAIN 398 457 LAMININ EGF-LIKE 3.
FT DOMAIN 458 509 LAMININ EGF-LIKE 4.
FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).
FT DOMAIN 541 771 LAMININ DOMAIN IV.
FT DOMAIN 773 820 LAMININ EGF-LIKE 6.
FT DOMAIN 821 866 LAMININ EGF-LIKE 7.
FT DOMAIN 867 916 LAMININ EGF-LIKE 8.
FT DOMAIN 917 975 LAMININ EGF-LIKE 9.
FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.
FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.

FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.
FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.
FT DOMAIN 1179 1397 DOMAIN II.
FT DOMAIN 1398 1430 DOMAIN ALPHA.
FT DOMAIN 1431 1786 DOMAIN I.
FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).
FT DOMAIN 1353 1388 COILED COIL (POTENTIAL).
FT DOMAIN 1442 1781 COILED COIL (POTENTIAL).
FT DISULFID 271 280 BY SIMILARITY.
FT DISULFID 273 298 BY SIMILARITY.
FT DISULFID 300 309 BY SIMILARITY.
FT DISULFID 312 332 BY SIMILARITY.
FT DISULFID 335 344 BY SIMILARITY.
FT DISULFID 337 362 BY SIMILARITY.
FT DISULFID 365 374 BY SIMILARITY.
FT DISULFID 377 395 BY SIMILARITY.
FT DISULFID 398 411 BY SIMILARITY.
FT DISULFID 400 426 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 440 455 BY SIMILARITY.
FT DISULFID 458 472 BY SIMILARITY.
FT DISULFID 460 479 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
FT DISULFID 493 507 BY SIMILARITY.
FT DISULFID 773 785 BY SIMILARITY.
FT DISULFID 775 792 BY SIMILARITY.
FT DISULFID 794 803 BY SIMILARITY.
FT DISULFID 806 818 BY SIMILARITY.
FT DISULFID 821 833 BY SIMILARITY.
FT DISULFID 823 840 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 864 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 869 883 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 898 914 BY SIMILARITY.
FT DISULFID 917 933 BY SIMILARITY.
FT DISULFID 919 944 BY SIMILARITY.
FT DISULFID 946 955 BY SIMILARITY.

Query Match 5.5%; Score 88; DB 1; Length 1786;
Best Local Similarity 18.4%; Pred. No. 10;
Matches 45; Conservative 36; Mismatches 115; Indels 48; Gaps 8;

Qy 82 PSGP-----ALQPPKITLSCYDCFSGSCNCSNNLRHELRKAKIACLDLR 132
Db 920 PDGPDGGRQFARSCYQDPVTIQLACVDPGYIGSCDDCASCYFONPSEVGSCQPCQCH 979

Qy 133 N-----PEAFQPKMKG-----GRLGAALKRHSKGCNCKRSCLKNY 168
Db 980 NNIDTDPDPCADKGTGRCLKCLYHTEGEHCQPCRFYGGDALRQDCRKCVCNLTGVQEH 1039

Qy 169 -----CECYEAKIMCSSICKIACKNYEESPERKMLMSTPHYNEPQDPFESSHYLSPA--K 221
Db 1040 CNGSDCQCDKATGQCLCLPNVIG-QNCDRCAPNTWQLASGTGCDPCNCAHSPGSCNE 1098

Qy 222 FSGPPKLRKRAQFSC-----ISWEVWEATCACLQAQEEAEQHCSPSLAEOMILE--E 274
Db 1099 FTGQCQCPMPGGRTCTSECQELFWGDDPVECRACDPRGIETPQCDQSTGQCVCVEGVE 1158

Qy 275 FGRC 278
Db 1159 GPRC 1162

RESULT 15
Y070_NPVAC STANDARD; PRT; 290 AA.
ID Y070_NPVAC
AC P41470;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 34.4 kDa protein in LEF3-IAP2 intergenic region.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:14:21 ; Search time 92.59 Seconds

(without alignments)
551.177 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 NVICQLKGAQMLCIDNCGA.....GRCLSQLIHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archesp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	%	Description
1	504	31.5	277	11 Q9D571
2	435	27.2	950	5 Q9V608
3	396	24.8	429	5 Q62295
4	396	24.8	435	5 Q95QD7
5	363	22.7	571	10 Q9SL70
6	357	22.3	603	10 Q9SZD1
7	253.5	15.9	356	10 Q9LW71
8	251.5	15.7	601	10 Q9CAV1
9	232.5	14.5	896	10 Q9ZS22
10	232	14.5	608	10 Q23333
11	223.5	14.0	609	10 Q9M679
12	222	13.9	526	10 Q9A12
13	222	13.9	695	10 Q9LE32
14	222	13.9	695	10 Q9L013
15	186	11.6	243	5 Q9VMQ3
16	180.5	11.3	593	10 Q9LUI5

17	159	9.9	553	10	Q94DS2
18	115	7.2	1737	4	Q75097
19	110.5	6.9	971	4	Q95980
20	103.5	6.5	643	11	Q9ET61
21	103.5	6.5	643	11	Q9JIZ6
22	100	6.3	773	5	O17514
23	100	6.3	773	5	O62335
24	99.5	6.2	1704	5	Q94446
25	99	6.2	2931	5	Q9W2C6
26	97.5	6.1	772	13	Q9P0U4
27	97	6.1	497	5	Q23460
28	97	6.1	984	5	Q9Y1P7
29	97	6.1	1106	5	Q9GU55
30	97	6.1	1987	5	Q9V841
31	96.5	6.0	4547	5	Q9W343
32	96	6.0	773	5	Q9V0R1
33	95.5	6.0	448	5	Q9NKE0
34	95.5	6.0	1035	5	Q9VJU8
35	95	5.9	1428	5	O44341
36	94.5	5.9	631	15	Q9IEZ0
37	94	5.9	716	5	Q9U2A6
38	94	5.9	1819	16	Q9SLV0
39	93.5	5.8	420	5	P91776
40	93.5	5.8	652	4	Q9NPY3
41	93.5	5.8	652	4	O00274
42	93.5	5.8	1981	5	Q9VVK7
43	93	5.8	761	5	O76733
44	93	5.8	4123	4	O75851
45	92.5	5.8	489	11	Q9D211

ALIGNMENTS

RESULT 1

Q9D571	Q9D571	PRELIMINARY;	PRT;	277 AA.
AC	Q9D571;			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,			
DE	CLONE:4930509C02, FULL INSERT SEQUENCE.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikolaide I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,			
RA	Hayashizaki Y.;			
RT	*Functional annotation of a full-length mouse cDNA collection.*;			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK015732; BAB29949.1; .			
SQ	SEQUENCE 277 AA; 28263 MW; 1731F517A3CB4D43 CRC64;			

Q94DS2	oryza sativ
O75097	homo sapien
Q95980	homo sapien
Q9ET61	rattus norv
Q9JIZ6	rattus norv
O17514	caenorhabdi
O62335	caenorhabdi
Q94446	chironomus
Q9W2C6	drosophila
Q9P0U4	ictalurus p
Q23460	caenorhabdi
Q9Y1P7	cryptospori
Q9GU55	drosophila
Q9V841	drosophila
Q9W343	drosophila
Q9VQR1	drosophila
Q9NKE0	drosophila
Q9VJU8	drosophila
O44341	halotis ru
Q9IEZ0	caprine art
Q9U2A6	caenorhabdi
Q9SLV0	helicobacte
P91776	pacifastacu
Q9NPY3	homo sapien
O00274	homo sapien
Q9VVK7	drosophila
O76733	drosophila
O75851	homo sapien
Q9D211	mus musculu

Query Match 31.5%; Score 504; DB 11; Length 277;
Best Local Similarity 100.0%; Pred. No. 5.3e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTCLKGGAGMCLINCAGARELKALHLLPQYDDQSSFFQSELPKMTTLVGRLLPVPK 60
|||||
Db 181 MVTCLKGGAGMCLINCAGARELKALHLLPQYDDQSSFFQSELPKMTTLVGRLLPVPK 240
|||||

QY 61 LNLITQVNGALPSAVNGAAPPSPALQGGPKITLSS 97
|||||
Db 241 LNLITQVNGALPSAVNGAAPPSPALQGGPKITLSS 277
|||||

RESULT 2
QYV608 PRELIMINARY; PRT; 950 AA.
ID Q9V608
AC Q9V608;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE CG6061 PROTEIN.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbavani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellia A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL: AF003818; RAF53365.1; FlyBase: FBgn0033846; CG6061.

SQL SEQUENCE 950 AA; 100021 MW; 627C3EA6B44A0A30 CRC64;
Query Match 27.2%; Score 435; DB 5; Length 950;
Best Local Similarity 38.4%; Pred. No. 6.9e-38;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

QY 31 QYDDQSSFFQSELPKMTTLVGRLLPVPKLNLIITQVNGALPSAVNGAAPPSPALQGP 90
|||
Db 681 QLPTEQSTPIKVEPKLPTLPVGVKANVPK--PLPEVLKPPATAAAGAVDPLGGMTSRR 738
|||

QY 91 PKITLSS-----YDCFSFGDFCNSCSC-----NNLRHELERFKAICACLDNRPEAFQPK 140
|||
Db 739 KHCNCSKSQCLKLYDCFANGFCQDCTCKDCFNLDYEVERAIRSCLDRNPSAFKPK 798
|||

QY 141 MGKGRGLGAALKRHSKCNCKRGCKNKCCEYAEKIMCSCICKIACKNVEESPCKMLM 200
|||
Db 799 ITAPNSGDMRL-HNKGCKNCKRGCKNKCCEYAEKIPCSICKKVCGRNWDKPD----- 852
|||

QY 201 STPHYMEPGDFESSHYLSPAKFSGPPKLRNKKQ-----AFSCISWEVVEATCACLAAQGE 255
|||
Db 853 -----VMDSLDGLMGVGGQKDK-AKKQLNENRANIVFTDDVIEATIMCMISRV 903
|||

QY 256 EAEQEHCSPLAEOMILEEFGRLCSOIL 283
|||
Db 904 MHEKQNAVVEDMEREYMEEMGESLTOII 931
|||

RESULT 3
ID Q62295 PRELIMINARY; PRT; 429 AA.
AC Q62295;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE JC8-6B PROTEIN.
GN JC8-6B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lightning J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology";
RL Science 282:2012-2018(1998).
DR EMBL: Z82274; CAB05228.1;
SQ SEQUENCE 429 AA; 48365 MW; 10601B22681C06D2 CRC64;

Query Match 24.8%; Score 396; DB 5; Length 429;
Best Local Similarity 38.9%; Pred. No. 4e-34;
Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

QY 98 YDCFSFGDFCNSCSC-----NNLRHELERFKAICACLDNRPEAFQPKMGKGRIGLGAALKR- 152
|||
Db 182 YDCFANGFCRDCNCKDCNHNIEYDSQRKAIQSLERNPNAFKPKIGIARGITDIER 241
|||

QY 153 -HSKGCNCKRGCKNKCCEYAEKIMCSCICKIACKN-----YEES-----PERKML 199
|||
Db 242 LHQKCHCKRCKNKCCEYAEKIPCTDRCKCKGCKGCTETRTMTYKNSGGAVSTNAL 301
|||

QY 200 MS-----TPHYMEPGDFESSHY-----LSPAKFSGPPKLRNKKRQAFSCISW-----EYV 243
|||
Db 302 MSLTNASSATPD-SGPGSVVTDHGDYEDMLSHKPKVEMDPRRF---PWYMTDEVV 357
|||

QY 244 EATCACLAAQGEA-----EAEHCSPLAEOMILEEFGRLCSOIL 283

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL078470; CAB43914.1; -

DR EMBL: AL161574; CAB79658.1; -

SQ SEQUENCE 603 AA; 64635 MW; 7FD1951AF818DE5 CRC64;

Query Match 22.3%; Score 357; DB 10; Length 603;

Best Local Similarity 35.2%; Pred. No. 9.9e-30;

Matches 88; Conservative 38; Mismatches 66; Indels 58; Gaps 12;

QY 36 SPSQELPKPWTTLVGRLLPVPKALNLTQVNGALPSAVNGA-----APFSPALQGP- 90

DB 60 SSVPTIRP-GMTIAGQVTVRPTLPMTATMSNPQSQIVNAPIRHPIPEPKARGPR 118

QY 91 PKT-----TLGYCDGFCSSGDFCNSC-----NNLRHELEFKAIK 127

DB 119 PNVGRDGTPOKKQCNKHSRCLKLYCECFASGTYCDGCNCVNCFNVDNEPARREAVE 178

QY 128 ACLDRNPEAFQPKMGK-----LGAALK--RHSKGCNCKRGCLKNYCEYEAKI 176

DB 179 ATLERNPEAFPRKIASPHGGGRKREDIGEVLLGKHNGCHCKGCLLKLYCECFQANI 238

QY 177 MCSICKCIACKNYSPERKMLMSTPH-----YME-----PGDFESSHYL-SPAKFS 223

DB 239 LCSSENCKLDCKNFEQSEERQALFHGEHNMAYLQOAAANAATGAVGSSGFAPSPA--- 295

QY 224 GPPKLRKRO 233

DB 296 --PKRRKQGE 303

RESULT 7

Q9LW71

ID Q9LW71 PRELIMINARY; PRT; 356 AA.

AC Q9LW71

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE GENOMIC DNA, CHROMOSOME 3, P1 CLONE: MSL1.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;

RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Nakamura Y.;

RA MEDLINE=20277480; PubMed=10819329;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence

RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC

RT clones."

RL DNA Res. 7:131-135(2000).

DR EMBL: AB012247; BAB02682.1; -

DR InterPro: IPR000345; CytC_heme_bind.

DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN1.

SQ SEQUENCE 356 AA; 39796 MW; 4FE0EA2D9DDBF58 CRC64;

Query Match 15.9%; Score 253.5; DB 10; Length 356;

Best Local Similarity 32.1%; Pred. No. 7.2e-19;

Matches 77; Conservative 30; Mismatches 90; Indels 43; Gaps 11;

QY 98 YDCDFSSGDFCNSCNNLRHELEFKAIKAC-----LDRNPEAFQPKMGKRLGAALRH 153

DB 79 YDCDFASVVCTDCDCVCHNNSEKCDAREAMVNLGRNPNFSEK-ALGSLTDNCKA 137

QY 154 -----SKGCNCKRSGCLKNYCEYEAKIMCSTICKCIACKNYEES--PERKMLMST 202

DB 138 APDTKPGLLSRGCKRTRCLKKYCECFQANLDCNCKNCNVSEAFQPPAFSAHNS 197

QY 203 PH-YMEPGDFESSHYLS-PAK-FSGP-----PKL--RKNRQAFSCISWEV 242

DB 198 PQYRRRRDRTELTEWNSCPAPLFSPDNISQNALGSPMSCSPALPYRKRLSMGYTSTLL 257

QY 243 VEA--TCACLLAQGEA-----EQHCSPSLAEPQMILEFGRLCSQILHIEFKSKLKTIE 295

DB 258 PDGLDGLCSLVAESATTTAEDQNRIFTKPDDBKEAIELSESSESRNVEEIQSGRLIE 317

RESULT 8

Q9CAV1

ID Q9CAV1 PRELIMINARY; PRT; 601 AA.

AC Q9CAV1

DT 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DE HYPOTHETICAL 66.8 KDA PROTEIN.

GN T9J14.20.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,

RA Farmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,

RA Delaney M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,

RA De Simone V., Choisine N., Artiguenave F., Robert C., Brottier P.,

RA Wincker P., Cattolico L., Weissbach J., Saurin W., Ouetier F.,

RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,

RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,

RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,

RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simonati B.,

RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,

RA Reichelt J., Scharfe M., Schoen O., Bagues M., Terol J., Climent J.,

RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,

RA Cooke R., Laude M., Berger-Llauro C., Purnelle B., Masuy D.,

RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,

RA Monfort A., Argirou A., Flores M., Liguori R., Vitale D.,

RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,

RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,

RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,

RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,

RA Pai G., Millischer J., Sellers P., Gill J.E., Feldblyum T.V.,

RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,

RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,

RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,

RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,

RA Watanabe A., Yamada M., Yasuda M., Tabata S.;

RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis

RT thaliana."

RL Nature 408:820-822(2000).

DR EMBL: AC009465; AAC51411.1; -

KW Hypothetical protein.

SQ SEQUENCE 601 AA; 66786 MW; 10848E970D81E022 CRC64;

Query Match 15.7%; Score 251.5; DB 10; Length 601;

Best Local Similarity 28.6%; Pred. No. 2.3e-18;

Matches 68; Conservative 28; Mismatches 59; Indels 83; Gaps 11;

QY 29 LPQYDQSSFPQ-----SELPKPMTTL-----VGRLLPVPKALNLTQVNGALSAV- 76

DB 304 IPLRSTNDLPIDSTISINKAPSPQNCIDTQKDTDEILPIPTIGLHL---NGFVNPSVS 360

QY 77 -----NGAAPP-----SGPALQPPKTLISG----- 97

Db 398 EVGEDASKTPASARHRGCKNCKSLKRYCEYQGVGCSNCRCEGCKNAFGRKDGSL 457
QY 190 YEESPERKMLMSTPHYMEPGDFSSHVLPSPAKFSGPPKLRKNRQ 233
Db 458 FEDEENET-----SGTPGKTKTOO 477

RESULT 12
Q94AI2 PRELIMINARY; PRT; 526 AA.
AC Q94AI2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE DNA BINDING PROTEIN
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Gibson H.A.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yu G., Yu S., Bowser L., Carninci P., Chen H.,
RA Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Natusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene MW123.15/AT3g22780 (GI:9279696).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046019; AAK76693.1;
SQ SEQUENCE 526 AA; 57886 MW; C6CD9684D417A9D9 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 526;
Best Local Similarity 31.8%; Pred. No. 3e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YDCFFSSGDFC-NSCSC-----NNLRHELEFRKAIKACLDNRNPAFQPKM----- 141
Db 245 YCECFAAGVYCIIEPCSCIDCFNKPHEETVLAIRKQIESRNPLAFAPKVI RNADSIMEAS 304
QY 142 GKGLGAAKLHSGCKNCKRGCKNCKYCEYQGVGCSNCRCEGCKTVNFGKDGSLVI 189
Db 305 DDASKTPASARHRGCKNCKNCKMKKYCEYQGVGCSNCRCEGCKTVNFGKDGSLVI 364
QY 190 -----YEESPE--RKMLMSTPHYME-----PGDFSSHVLPSPAK--FSGPPKLRKNR 232
Db 365 MESKLEENQETYEKRIAKIQHNVEVSKEVEQNPSDQSTPLPPYRLVHVHQPFLSKNR 423

RESULT 13
Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CXC DOMAIN PROTEIN TSOL (PUTATIVE DNA BINDING PROTEIN).
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LER;
RX MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;

RT TSOL is a novel protein that modulates cytokinesis and cell expansion
in Arabidopsis.";
RL Development 127:2219-2226(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song J.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TSOL, an
Arabidopsis gene with cysteine-rich repeats.";
RL Development 127:2207-2217(2000).
DR EMBL: AF204059; AAF69124.1;
DR EMBL: AF206324; AAF27433.1;
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA0850F6B5D4 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 695;
Best Local Similarity 31.8%; Pred. No. 4.2e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YDCFFSSGDFC-NSCSC-----NNLRHELEFRKAIKACLDNRNPAFQPKM----- 141
Db 414 YCECFAAGVYCIIEPCSCIDCFNKPHEETVLAIRKQIESRNPLAFAPKVI RNADSIMEAS 473
QY 142 GKGLGAAKLHSGCKNCKRGCKNCKYCEYQGVGCSNCRCEGCKTVNFGKDGSLVI 533
Db 474 DDASKTPASARHRGCKNCKNCKMKKYCEYQGVGCSNCRCEGCKTVNFGKDGSLVI 592
QY 190 -----YEESPE--RKMLMSTPHYME-----PGDFSSHVLPSPAK--FSGPPKLRKNR 232
Db 534 MESKLEENQETYEKRIAKIQHNVEVSKEVEQNPSDQSTPLPPYRLVHVHQPFLSKNR 592

RESULT 14
Q9LUI3 PRELIMINARY; PRT; 695 AA.
AC Q9LUI3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE DNA BINDING PROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty P1 and TAC
clones.";
RL DNA Res. 7:131-135(2000).
DR EMBL: AB022223; BAB01253.1;
SQ SEQUENCE 695 AA; 76276 MW; 34BBA08450F6BCE1 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 695;
Best Local Similarity 31.8%; Pred. No. 4.2e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YDCFFSSGDFC-NSCSC-----NNLRHELEFRKAIKACLDNRNPAFQPKM----- 141
Db 414 YCECFAAGVYCIIEPCSCIDCFNKPHEETVLAIRKQIESRNPLAFAPKVI RNADSIMEAS 473
QY 142 GKGLGAAKLHSGCKNCKRGCKNCKYCEYQGVGCSNCRCEGCKTVNFGKDGSLVI 592

DB 474 DASKTPASARHKGCGNCKNCKMCKYCEYGGVGCNMCRCEGCTNFGRRGDSLLVI 533
 QY 190 ----YEEPE--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPKLRKNR 232
 DB 534 MESKLEENQETYEKRIAKTOHNVENSKEVEQNPPSSDQPTPLPPYRHLVYVHOPFLSKNR 592

RESULT 15
 Q9VMQ3 PRELIMINARY; PRT: 243 AA.
 AC Q9VMQ3;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE CG14016 PROTEIN.
 GN CG14016.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Brill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AF003610; AAF52259.1; --
 DR HSP: F10414; 1BBG.
 DR FlyBase: FBgn0031715; CGI4016.
 SQ SEQUENCE 243 AA; 26179 MW; B3E11FA1D2605934 CRC64;

Query Match 11.6%; Score 186; DB 5; Length 243;
 Best Local Similarity 25.4%; Pred. No. 8.4e-12;
 Matches 59; Conservative 26; Mismatches 61; Indels 86; Gaps 8;

QY 132 RNPEAFQPMKGRGLGAALRHSGKNCGRSGCLKNVCYCEYEAkimcsstckciacknye 191
 DB 7 RSVDRKADGKGGKQ--CAGGV---KGCCCRSRQCIKNYCDYQSMaictkfcrcvgrnte 62
 QY 192 -----ESPER-----KMLM 200
 DB 63 VRELVDPNVSAKNSSAVKQKAAAKAAAGIDVQKALQVAASTLALPGKALM 122
 QY 201 STPHYM-----EPGDFESSHYLSPAKFSGP-----PKLRKN 231
 DB 123 TPKYTLVAGKP--PMASHH-INDIPISRTAATAATPARAVKQPAEPPMPVNLIIPIVRHDD 180
 QY 232 ROAFSCISWEVVEATCACLLAOGEEAEQEHCSPLAEOMTLEEFGRCLSOIL 283
 DB 181 RDRNLFVQPVNAALLECMLEIQATEAEQGLNQLVQCVLVEEFMRGYNIL 232

Search completed: July 11, 2002, 08:22:18
 Job time: 477 sec

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	93.5	5.8	652	2	US-08-751-305-2	Sequence 2, Appli	
2	87.5	5.5	610	1	US-08-365-470-3	Sequence 3, Appli	
3	87.5	5.5	610	3	US-09-203-668-19	Sequence 19, Appl	
4	87.5	5.5	610	4	US-09-003-490A-89	Sequence 89, Appl	
5	86	5.4	418	4	US-08-795-430-13	Sequence 13, Appl	
6	86	5.4	610	6	5217870-2	Patent No. 5217870	
7	86	5.4	4	830	1	Sequence 4, Appli	
8	85.5	5.3	769	2	US-08-110-158-4	Sequence 1, Appli	
9	85.5	5.3	769	2	US-08-789-078-1	Sequence 1, Appli	
10	85.5	5.3	769	2	US-08-752-633-1	Sequence 1, Appli	
11	85.5	5.3	769	2	US-08-476-062A-45	Sequence 45, Appl	
12	85.5	5.3	769	2	US-07-728-215-31	Sequence 31, Appl	
13	85.5	5.3	769	4	US-08-938-085A-31	Sequence 31, Appl	
14	85.5	5.3	769	5	PCT-US95-04886-1	Sequence 1, Appli	
15	85.5	5.3	769	5	PCT-US96-01314-45	Sequence 45, Appl	
16	85	5.3	676	3	US-08-630-172-10	Sequence 10, Appl	
17	84.5	5.3	696	3	US-09-375-419-10	Sequence 10, Appl	
18	84.5	5.3	696	4	US-08-899-437-23	Sequence 23, Appl	
19	84.5	5.3	720	3	US-09-126-121-23	Sequence 23, Appl	
20	84.5	5.3	720	3	US-08-899-437-6	Sequence 6, Appli	
21	83.5	5.2	350	2	US-09-126-121-6	Sequence 6, Appli	
22	83.5	5.2	350	2	US-08-999-811-4	Sequence 4, Appli	
23	83.5	5.2	350	3	US-08-824-996-2	Sequence 2, Appli	
24	83.5	5.2	350	4	US-09-042-105-4	Sequence 4, Appli	
25	83.5	5.2	350	4	US-08-510-133A-33	Sequence 33, Appl	
26	83.5	5.2	419	2	US-08-583-895-33	Sequence 33, Appl	
27	83.5	5.2	419	3	US-08-999-811-2	Sequence 2, Appli	
28	83.5	5.2	419	3	US-09-042-105-2	Sequence 2, Appli	

```
Query Match          5.8%; Score 93.5; DB 2; Length 652;
Best Local Similarity 27.0%; Pred. No. 0.57;
Matches 30; Conservative 11; Mismatches 51; Indels 19; Gaps 4;
```

QY 96 SGYC--DCSSGGFCNSCNNRHELERFKAICACLDNRPEAFQPMKGRGLCAALRH 153
 :| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 268 NGGCHODCFEGGDGSFLCGC---RPGFRLDDLVTCAARNPCSSSPCRG----- 313
 :| | | | | | | | | | | | | | | | | | | | | | | | | |

QY 154 SKGCNCKRRSGCLKNY-CECYEAIMCSSICKCIACKNYEESPERKMLMSTP 203
 :| | | | | | | | | | | | | | | | | | | | | | | | | |

Db 314 --GATCVLGPBGKNTCTCPQGYQLDSSOLDVVDVDECDSPCAQECVNTVP 362

RESULT 2

US-08-365-470-3
; Sequence 3, Application US/08365470
; Patent No. 5632991
; GENERAL INFORMATION:
; APPLICANT: Gimbrone, Jr., Michael A.
; TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,470
; FILING DATE: herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,510
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/850,802
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Markowicz, Karen R.
; REGISTRATION NUMBER: 36,351
; REFERENCE/DOCKET NUMBER: 0627.1350003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-365-470-3

Query Match 5.5%; Score 87.5; DB 1; Length 610;
Best Local Similarity 21.6%; Pred. No. 2;
Matches 74; Conservative 33; Mismatches 119; Indels 117; Gaps 20;

QY 15 IDNCGARE-----LKALHLLPOYDDQS-----SFPQSELKPKMTTL-----VGR-LLPVPA 59
Db 177 IVNCTALESPEHGLVCSHPLGNFSYNSCSISCDRGYLPSSMETMQCMSSGEWSADIPA 236
QY 60 ----KLNLITQVDNGALPSAVNGAAPP-----SGPALQGPPIKTLISGYDCDFSSG 105
Db 237 CNVVECDAVTNPANGFVECFQNGSPFWNTCTFDCGEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLRHELERFKAIAKACLDNRNPEAFQPMKGRIG-----AAKLHRSKG 156
Db 291 NWDNEKPTCAVTCRAVR-----QPQNGSVRCSHSPAGEFTFKSS 330
QY 157 CN--CKRSGLKNYCEYEAKIMCSS-----ICKCIACKNYEESPERKMLMSTPHY 205
Db 331 CNFTCEEGFMLOG-----PAQVECTTQGTQIPVCEAFQCTAL-SNPER-----GY 377
QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLKRNQAFSCISWEVVEATCACLIAQ----- 253
Db 378 MNCLPSASGSFRYSGSCFSCFCEQGFVLKSKRLQCGTGEWDNEKPTCEAVRCDVAHQPP 437

QY 254 -----GEEAQEHCSPLAEQMILEEFG-----RCLSQ 281
Db 438 KGLVRCAHSPIGEFTYKSSCAFSCEEG--FELYGSTQLECTSQ 478

RESULT 3

US-09-209-668-19
; Sequence 19, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-668-19

Query Match 5.5%; Score 87.5; DB 3; Length 610;
Best Local Similarity 21.6%; Pred. No. 2;
Matches 74; Conservative 33; Mismatches 119; Indels 117; Gaps 20;

QY 15 IDNCGARE-----LKALHLLPOYDDQS-----SFPQSELKPKMTTL-----VGR-LLPVPA 59
Db 177 IVNCTALESPEHGLVCSHPLGNFSYNSCSISCDRGYLPSSMETMQCMSSGEWSADIPA 236
QY 60 ----KLNLITQVDNGALPSAVNGAAPP-----SGPALQGPPIKTLISGYDCDFSSG 105
Db 237 CNVVECDAVTNPANGFVECFQNGSPFWNTCTFDCGEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLRHELERFKAIAKACLDNRNPEAFQPMKGRIG-----AAKLHRSKG 156
Db 291 NWDNEKPTCAVTCRAVR-----QPQNGSVRCSHSPAGEFTFKSS 330
QY 157 CN--CKRSGLKNYCEYEAKIMCSS-----ICKCIACKNYEESPERKMLMSTPHY 205
Db 331 CNFTCEEGFMLOG-----PAQVECTTQGTQIPVCEAFQCTAL-SNPER-----GY 377
QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLKRNQAFSCISWEVVEATCACLIAQ----- 253
Db 378 MNCLPSASGSFRYSGSCFSCFCEQGFVLKSKRLQCGTGEWDNEKPTCEAVRCDVAHQPP 437
QY 254 -----GEEAQEHCSPLAEQMILEEFG-----RCLSQ 281
Db 438 KGLVRCAHSPIGEFTYKSSCAFSCEEG--FELYGSTQLECTSQ 478

RESULT 4

US-09-009-490A-89
; Sequence 89, Application US/09009490A
; Patent No. 6300491
; GENERAL INFORMATION:
; APPLICANT: Bennett and Mirabelli
; TITLE OF INVENTION: Oligonucleotide Modulation
; TITLE OF INVENTION: of Cell Adhesion
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Jane Massey Licata
; STREET: 66 East Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE

Db 280 GPNKELDEETCQCVCKGVRPISCGPHKELDRASCQCMCKNKLPSGCCGNKEFDEEKCQ 339
QY 117 -----RHLEERFKAKACLDNRPEAFQPMKGRGLGAALKRHSKGNCKRSGLKN 167
Db 340 CVCKKTCPRHPLNPAKCECTESNKFL-----KG-----KRFHQVCYVRPCTVR 390
QY 168 YCEYEAKIMCISICKI 185
Db 391 TKRCDAFLAEEVCRV 408
RESULT 6
5217870-2
; Patent No. 5217870
; APPLICANT: HESSON, CATHERINE A.; LOBB, ROY R.; GOELZ, SUSAN E.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST CDX
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/345,151
; FILING DATE: 28-APR-1989
; SEQ ID NO: 2:
; LENGTH: 610
5217870-2

Query Match 5.4%; Score 86; DB 6; Length 610;
Best Local Similarity 22.2%; Pred. No. 2,9;
Matches 64; Conservative 29; Mismatches 101; Indels 94; Gaps 17;
QY 15 IDNCGARE-----LKAHLLOYDQDS-----SFPQSELPKPMTTL-----VGR-LLPVPA 59
Db 177 IVNCTALESPEHSLVCSHPLGNFSYSSCSISCDRGLPSPMETQCMSSGEWSAPIPA 236
QY 60 -----KLNLIQVONGALPSAVNGAAPP-----SGPALOGPPKITLSGYDCDFSSG 105
Db 237 CNVVECDAVTNPANGFVECFQNGSPFWNTCTFDCBEGFELMGAOSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLRHELERFKAKACLDNRPEAFQPMKGRGLG-----AAKLRHSG 156
Db 291 NWDNEKPTCKAVTCRAVR-----QPNGSVRCSSHSPAGEFTFKSS 330
QY 157 CN-----CKRSGLKNYCEYEAKIMCSC-----ICKCIACKNYEESPERKMLMSTPHY 205
Db 331 CNFTCEGFMLOG-----PAQVECTQGWTOQIPVCEAFQCTAL-SNPER-----GY 377
QY 206 ME-----PGDFESSHYLSPAKFSPPP-----KLKRNQAFSCISWEVVEATC 247
Db 378 MNCLPSASGSFRYGSCEFSCEQGFVLKSKRLQCGPTGEMDNEKPTC 425

RESULT 7
US-08-110-158-4
; Sequence 4, Application US/08110158
; Patent No. 5605821
; GENERAL INFORMATION:
; APPLICANT: McEever, Rodger P.
; APPLICANT: Pan, Junliang
; TITLE OF INVENTION: Expression Control Sequences of the
; TITLE OF INVENTION: P-Selectin Gene
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/110,158
; FILING DATE: 19930820
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/320,408
; FILING DATE: 08-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-815-6508
; TELEFAX: (404)-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 830 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-110-158-4

Query Match 5.4%; Score 86; DB 1; Length 830;
Best Local Similarity 22.4%; Pred. No. 4,3;
Matches 46; Conservative 24; Mismatches 79; Indels 56; Gaps 12;
QY 84 GPALOGPPKITLSGYDCDFSSGDF-----CNSCSCNNLRHELERFKAKACLDNRPEAF 137
Db 233 GYQVNGPSKL-----ECLASGIWTKNPQCLAAQCPPLKIP-ERGNMI--CL-HSAKAF 282
QY 138 QPMKGRGLGAALKRHSKGNCKRSGLKNYCEYEAKIMCSC-----ICKCIACK 188
Db 283 Q-----HQSOSCS-----FSCGEGFALVGPVQCTASGVMTAPAPYCKAVQCQ 325
QY 189 NYEESPERKMLMSTPHYMPGDFESSHYLSPAKFSPPKLR-KNRQAFSCIS---WEVVE 244
Db 326 HLEAPSEGTMDCVHP-----LTAAYGSSCKFECPGYVRGLDMLRCLDSDGHSNAPL 378
QY 245 ATCACLQAQGEAE-----QEHCSPL 266
Db 379 PTCEAISCEPLESPVHGMDSPSL 403

RESULT 8
US-08-789-078-1
; Sequence 1, Application US/08789078
; Patent No. 5843885
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbets, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,078
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,513

FILING DATE: 19-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: (816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label= cyto
OTHER INFORMATION: /note= "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
AUTHORS: Power,
TITLE: LFA-1 Amino acid sequence (B2) (from human
TITLE: tonsil)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
US-08-789-078-1

Query Match 5.3%; Score 85.5; DB 2: Length 769;
Best Local Similarity 23.4%; Pred. No. 4.4;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps 10;

Qy 90 PPKITLSCYDCGFS-----SGD-----FCNSCSN-----NLRIELERFKAIKAC 129
| | | | | : | | | | : | | | : | | | : |
Db 525 PGLIYQYCECDTINCERYNGVCGPGRLGICFGKCRCHPFGSGACQGER--TTEGC 582

Qy 130 LDRNPEAFQPKMGKRLGAALKRHSKGNCKRSGLKNVCECYEAKIM-----CSSI 181
| | | | | : | | | | | : | | | | : | | | : |
Db 583 L--NPRRVECS-GRGR-----CRC-----NVCECHSGVQLPLCQECGCFSP 621
| | | | | : | | | | | : | | | | : | | | : |
Qy 182 C-----KCIACKNYEESP 194
| | | | | : | | | | | : | | | | : | | | : |
Db 622 CGKYISCAECLKFEKGP 638
| | | | | : | | | | | : | | | | : | | | : |
RESULT 9:
US-08-752-633-1
; Sequence 1, Application US/08752633
; Patent No. 5863889
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Slahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752.633
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: (816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..22
; OTHER INFORMATION: /label= signal
; OTHER INFORMATION: /note= "signal sequence"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 449..496
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region


```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

, COMPUTER: IBM PS/2 Model 502 or 55SX
, OPERATING SYSTEM: MS-DOS (Version 5.0)
, SOFTWARE: WordPerfect (Version 5.1)
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: PCT/US96/01314
, FILING DATE: 30-JAN-96
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 08/380,167
, FILING DATE: 30-JAN-95
, ATTORNEY/AGENT INFORMATION:
, NAME: John W. Freeman
, REGISTRATION NUMBER: 29,066
, REFERENCE/DOCKET NUMBER: 00786/267001
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (617) 542-5070
, TELEFAX: (617) 542-8906
, TELEX: 200154
, INFORMATION FOR SEQ ID NO: 45:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 769
, TYPE: amino acid
, STRANDEDNESS:
, TOPOLOGY: linear
, PCT-US96-01314-45

```

	Query Match	5.3%;	Score 85.5;	DB 5;	Length 769;	
	Best Local Similarity	23.4%;	Pred. No. 4.4;			
	Matches	32;	Conservative	13;	Mismatches	37; Indels 55; Gaps 10;
QY	90	PKKITISGYGDCDFS-----SGD-----FCNSCSN----	NURHELERPKATKAC	129		
		: :	: :	:		
DB	525	PGKLIGYQCECDTINCERYNGVCGPGRGLFCGRCRCHPFEGSACQCR--TTTEGC	582			
		: :	: :	:		
QY	130	LDRNPFAFPKMKGRLGAALKRHSGKCNCKRSGLKNKYCECYEAKITM-----CSSI	181			
		: :	: :	:		
DB	583	L--NPRRVES--GRGR-----CRC-----NVCEHSYGYPQLCOECPCGPSP	621			
		: :	: :	:		

Qy	90	PKITUSGYCDGFS-----SGD-----FONSCSN-----NLRHELFRAIKAC	129
Db	525	PGKLIYQYCECDTINCERYNGQVGGPGRGLFCGKCRCHFGESACQCR--TTEGC	582
Qy	130	LDRNPEAFOPKNGKRLGAALKRHSKGCKNSGCLKNYCEYEAKIM-----CSSI	181
Db	583	L--NPRRVECS-GRGR-----CRC-----NVCECHSGVQLPLCQCEGCGPSP	621
Qy	182	C-----KCIACKNYEESP	194
Db	622	CGKIYSCAECLXFERGP	638

```

RESULT 14
PCT-US96-01114-45
; Sequence 45, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaout
; TITLE OF INVENTION: METHODS FOR ID
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; NUMBER OF INVENTION: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44

```

RESULT 15.

US-08-630-172-10

; Sequence 10, Application US/08630172

; Patent No. 6060054

; GENERAL INFORMATION:

; APPLICANT: Scaerz, Uwe

; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T

; TITLE OF INVENTION: LYMPHOCYTE VETO

; NUMBER OF SEQUENCES: 41

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross & McIntosh

; STREET: 1700 Lincoln Street, 35th Floor

; CITY: Denver

; STATE: Colorado

; COUNTRY: U.S.

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/630.172

; FILING DATE:

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Connell, Gary J.

; REGISTRATION NUMBER: 32,020

; REFERENCE/DOCKET NUMBER: 2879-36

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 863-9700
 ; TELEFAX: (303) 863-0223
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 676 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ;
 US-08-630-172-10

Query Match 5.3%; Score 85; DB 3; Length 676;
 Best Local Similarity 22.1%; Pred. No. 4.1;
 Matches 30; Conservative 13; Mismatches 39; Indels 54; Gaps 9;

Qy	90	PPKITLSGYCDGFS-----SGD-----FCNSCSN-----NLRHELERFKAIKAC	129
Db	503	PKLIYGYCECDTINCERYNGVCGPGRLCGKRCRHPGFGSACQGER--TTEGC	560
Qy	130	LDRNPEAFQPKMGKGLGAALKRHSKGCNCKRSGLKNYCECYEAKIN-----C	178
Db	561	L--NPRRVES-GRGR-----CRC-----NWCECHSGYQLPLCQECGCPSC	599
Qy	179	SSICKCIACKNYEESP	194
Db	600	GKYISCAECLKFEKGP	615

Search completed: July 11, 2002, 08:16:00
 Job time: 144 sec


```

QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKTLGYCDGFCSSGDFCNSCNCNNLRHEL 120
Db 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKTLGYCDGFCSSGDFCNSCNCNNLRHEL 120
QY 121 ERFAIKACLDNRNPEAFQPMKGRGLGAALRHSGKNCNCRSGCLKNYCEYAKIMCSS 180
Db 121 ERFAIKACLDNRNPEAFQPMKGRGLGAALRHSGKNCNCRSGCLKNYCEYAKIMCSS 180
QY 181 IKCKIACKNYESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRKNROAFSCISW 240
Db 181 IKCKIACKNYESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRKNROAFSCISW 240
QY 241 EVBEATCACLLAQGEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295
Db 241 EVBEATCACLLAQGEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295

```

RESULT 2

```

US-09-743-237-24
; Sequence 24, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743, 237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-743-237-24

```

```

Query Match 100.0%; Score 1599; DB 21; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.3e-143;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MVICQLKGGAGMCLIDNCGARELKALHLLPOYDDQSSFFQSELPKPMPTTLVGRLLPVPK 60
Db 1 MVICQLKGGAGMCLIDNCGARELKALHLLPOYDDQSSFFQSELPKPMPTTLVGRLLPVPK 60
QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKTLGYCDGFCSSGDFCNSCNCNNLRHEL 120
Db 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKTLGYCDGFCSSGDFCNSCNCNNLRHEL 120
QY 121 ERFAIKACLDNRNPEAFQPMKGRGLGAALRHSGKNCNCRSGCLKNYCEYAKIMCSS 180
Db 121 ERFAIKACLDNRNPEAFQPMKGRGLGAALRHSGKNCNCRSGCLKNYCEYAKIMCSS 180
QY 181 IKCKIACKNYESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRKNROAFSCISW 240
Db 181 IKCKIACKNYESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRKNROAFSCISW 240
QY 241 EVBEATCACLLAQGEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295
Db 241 EVBEATCACLLAQGEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295

```

RESULT 3

```

US-09-743-237-5
; Sequence 5, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.

```

```

; APPLICANT: MITSUI, YOUNI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743, 237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-743-237-5

```

```

Query Match 78.2%; Score 1251; DB 21; Length 299;
Best Local Similarity 76.3%; Pred. No. 2.2e-110;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

```

```

QY 1 MVICQLKGGAGMCLIDNCGARELKALHLLPOYDDQSSFFQSELPKPMPTTLVGRLLPVPK 60
Db 1 MVICQLKGGAGMCLIDNCGARELKALHLLPOYDDQSSFFQSELPKPMPTTLVGRLLPVPK 60
QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKTLGYCDGFCSSGDFCNSCNCNNLRHEL 116
Db 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTKTLGYCDGFCSSGDFCNSCNCNNLRHEL 120
QY 117 RHELEFRKAIKACLDNRNPEAFQPMKGRGLGAALRHSGKNCNCRSGCLKNYCEYAKI 176
Db 121 HHDIERFKAIKACLDNRNPEAFQPMKGRGLGAALRHSGKNCNCRSGCLKNYCEYAKI 180
QY 177 MCSICKIACKNYESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRKNROAFS 236
Db 181 MCSICKIACKNYESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRKNROAFS 240
QY 237 CISWEVVEATCACLLAQGEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295
Db 241 CISWEVVEATCACLLAQGEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 299

```

RESULT 4

```

PCT-US01-08656-10693
; Sequence 10693, Application PC/TUS0108656
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-066
; CURRENT APPLICATION NUMBER: PCT/US01/08656
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 09/522,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: 09/770,160
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 10994
; SOFTWARE: Custom
; SEQ ID NO 10693
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08656-10693

```

```

Query Match 31.6%; Score 505.5; DB 1; Length 223;
Best Local Similarity 52.4%; Pred. No. 2.1e-39;
Matches 97; Conservative 22; Mismatches 35; Indels 31; Gaps 4;

```

```

QY 44 PKPMTTLVGRLLPVPKLNLTQVDNGALPSAVNGAAPPSPALQGPPTKTLG----- 97
Db 1 PAPCTGNVGVAV-LPAQ--YVTLQSSSYVSIASNSFTCTGTGTQARLPFGIIPSES 57
QY 98 -----YCDFSSGDFCNSCSC-----NLRHELEFRKAIKACLDNRPE 135

```



```

Db      58  ASRPRKPCNCTKSLSCLKLYCDFANGFCNCCNCTNCTNCTNNLEHENERQKAIKACLDNRPE 117
Qy      136  AFOPKMGKGRGLGAALKLRHSKGCNCKRSGCLKKNYCECYEARIMCSSSTCKCKIACKNYEESPE 195
      ||:||||: | : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      118  AFKPKIGKKGEGESDRRHSKGCNCKRSGCLKKNYCECYEARIMCSSSTCKCKIGCKNFEEESPE 177
Qy      196  RKMLM 200
      || ||
Db      178  RKTLM 182

RESULT 5
US-60-161-932-1700
; Sequence 1700, Application US/60161932
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic
; TITLE OF INVENTION: Acid Sequences, Systems Containin the Nucleic Acid Sequences
; TITLE OF INVENTION: and Uses Thereof
; FILE REFERENCE: CL000122
; CURRENT APPLICATION NUMBER: US/60/161,932
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 2626
; SOFTWARE: FastSEQ for Windows version 4.0
; SEQ ID NO 1700
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Drosophila
US-60-161-932-1700

```

```

Query Match          27.2%   Score 435;   DB 26;   Length 403;
Best Local Similarity 38.4%;   Pred. No. 2.6e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

QY  31 QYDDQSFFQSESLPKPMTTLVGRLLPVPKAKUNLITQVDNGALPSAVNGAASPSPALGGP 90
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   134 QLPTEQSTPIKVEFKLPTLPVGVKANYPAK--PLFEVLKPPATAAGAVDPLGWTSSR 191
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  91 PKITLSG-----YDCDFSGDFCNSCS-----NNLRHELRFKAIKACLDRNPEAFQPK 140
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   192 KHCNCSKSQLKLYCDCFANGFQCQDCTCKCFNNLDYEVRERAIRSCLDRNPSAFKPK 251
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  141 MGKRLGAALKRLRHSKGNCRSGCLKNYCECYEAKIMCSSICKCIACKNYEESPERKMLM 200
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   252 ITAPNSGDMRL-HNKGNCNCRSGCLKNYCECYEAKIPCCSSICKCVGCRNMDRPD----- 305
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  201 STPHMEPGDPESHYLSPAKFSPPKLKNRQ-----AFSCISWEVVEATCACLLAGE 255
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   306 -----VMDSDLDGLMGVGBQKDK-AKNKQLNENRANITYFTDDVIEATIMCMISRV 356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  256 EAEQHCSPSLAEQIMLEEFGRCLSQIL 283
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   357 MHEKQNAVEDMERVMEEMGESITQII 384
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-60-167-217-12948
; Sequence 12948, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12948
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Drosophila

```

```

US-60-167-217-12948

Query Match          27.2%;   Score 435;   DB 26;   Length 890;
Best Local Similarity 38.4%;   Pred. No. 7 8e-32;
Matches -103;   Conservative 40;   Mismatches 93;   Indels 32;   Gaps

Qy  31 QYDQSSFPQSELPKPMWTLVGRLLPVPAKLNLTQVDNGALPSAVNGAAPSPGALQGP 1400
Db  621 QLPTETQSTPIKVEPKLTPLPQGVKANYPAK--PLFEVHLKPPATAAAGAVDPLGGMTSRR 678
Qy  91 PKITLSG-----YDCDFSSGDFCNSCSC-----NNLRHELERFKAIKACLDNRPEAFQPK 1400
Db  679 KHCNCSKSOCLKLYCDCFANGFCQDCTCKDFNNLDYEVERERAIRSCLDRNPSAFKPK 7398
Qy  141 MGKGRGLGAALKRHSGKNCKRSGCLKNKYCEYEAKIMCSTICKCIACKNYEESPERKMLM 2000
Db  739 ITAPNSGDMRL--HNKGNCNCRSGCLKNKYCEYEAKIPCSSTICKCVGCRNMEDRPD----- 792
Qy  201 STPHMEPGDPFESSHYLSPAKFSGPPKLNRNQ-----AFSCTSVEVVEATCACLLAQGE 2555
Db  793 -----VMDMSLDGLMGVEGQKDK-AKNKQLNENRANIYFTDDVEATIMCMISRIV 843
Qy  256 EAEQHCSPSLAEQMIILEEFGRCISQIL 283
Db  844 MHEKQNAVDEMERVEEMEGESITQII 871

```

```

RESULT 7
US-09-614-150-12897
; Sequence 12897, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12897
; LENGTH: 950
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150-12897

```

	Query Match	27.2%	Score 435;	DB 20;	Length 950;
Best Local Similarity	38.4%;	Pred. No. 8.5e-32;			
Matches 103;	Conservative 40;	Mismatches 93;	Indels 32;	Gaps 7;	
Qy 31	QYDQSSPPSELKPKMTTLVGRLLPVPVPAKLNLTIVDNGALPSAVNGAAFPSPALQGP	90			
DB 681	OLPTPEOSTPIKVEPKLPTLPQGVKANVPK--PLFVFLKPPKATAAAGAVPLDGLGWTSSR	738			

Qy	91	PKITLSG-----YDCPSSGDFCNSCS-	----	NNLRHLEFRFAIKACLDNRNPEATQPK	140
Db		: : : :			
Db	739	KHCNCRSQCILKLYCDCFANGFECODCTCKCFNNLDYEVRERAIRSCLDRNP	SAPFKPK	798	
Qy	141	MKGRLGAALKRHSGKNCRSGCLKNYCECYEAKIMCSSICKIACKNYEESPERKMLM	200		
Db		: : : : : : : : : : : : : : : : : :			
Db	799	ITAPNSGDMRL-HNKGNCKRSGCLKNYCECYEAKIPCCSIKCVGCGRNMEDRPD	-----	852	
Qy	201	STPHMEPGDPFESSHYLSPAKFSPPKLKRNRQ	-----AFSCISWEVVEATCACLLAQGE	255	
Db		: : : : : : : : : : : : : : : : :			
Db	853	-----VDMSLDGLMGVBGGKKDK-AKNQLNENRANIYFTDDVIETATMCMI	SIRIV	903	
Qy	256	EAEQHCSPSLAEOQMILEEFGRCLSQIL	283		
Db		: : : : : : : : : :			
Db	904	MHEQNVAVEDMEREVMEEMGESLTQII	931		
 RESULT 8					
US-60-191-637-12935					
; Sequence 12935, Application US/60191637					
; GENERAL INFORMATION:					
; APPLICANT: Venter, J. Craig					
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING					
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND					
; TITLE OF INVENTION: USES THEREOF					
; FILE REFERENCE: CL000392					
; CURRENT APPLICATION NUMBER: US/60/191,637					
; CURRENT FILING DATE: 2000-03-23					
; NUMBER OF SEQ ID NOS: 42660					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 12935					
; LENGTH: 950					
; TYPE: PRT					
; ORGANISM: DROSOPHILA					
US-60-191-637-12935					
 Query Match 27.2%; Score 435; DB 26; Length 950;					
Best Local Similarity 38.4%; Pred. No. 8,Se-32;					
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;					
Qy	31	OYDQSSFPQSSELPKPMTTVLGRLLPVPAKLNLIQTVDNGALPSAVNGAAPFGPALQGP	90		
Db		: : : :			
Db	681	QLPTREQSTPIKVEKPLTLPBGKANVPAAK--PLFEVLKPAPTAAAAGAVDPLGMTSR	738		
Qy	91	PKITLSG-----YDCPSSGDFCNSCS-	----	NNLRHLEFRFAIKACLDNRNPEATQPK	140
Db		: : : :			
Db	739	KHCNCRSQCILKLYCDCFANGFECODCTCKCFNNLDYEVRERAIRSCLDRNP	SAPFKPK	798	
Qy	141	MKGRLGAALKRHSGKNCRSGCLKNYCECYEAKIMCSSICKIACKNYEESPERKMLM	200		
Db		: : : : : : : : : : : : : : : : : :			
Db	799	ITAPNSGDMRL-HNKGNCKRSGCLKNYCECYEAKIPCCSIKCVGCGRNMEDRPD	-----	852	
Qy	201	STPHMEPGDPFESSHYLSPAKFSPPKLKRNRQ	-----AFSCISWEVVEATCACLLAQGE	255	
Db		: : : : : : : : : : : : : : : : :			
Db	853	-----VDMSLDGLMGVBGGKKDK-AKNQLNENRANIYFTDDVIETATMCMI	SIRIV	903	
Qy	256	EAEQHCSPSLAEOQMILEEFGRCLSQIL	283		
Db		: : : : : : : : : :			
Db	904	MHEQNVAVEDMEREVMEEMGESLTQII	931		
 RESULT 9					
US-09-087-136-13					
; Sequence 13, Application US/09087136					
; GENERAL INFORMATION:					
; APPLICANT: H. Robert Horvitz					
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS					
; FILE REFERENCE: 01997/202002					
; CURRENT APPLICATION NUMBER: US/09/087,136					
; CURRENT FILING DATE: 1998-05-28					
; EARLIER APPLICATION NUMBER: 60/047,996					
; EARLIER FILING DATE: 1997-05-28					

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:13:36 ; Search time 30.19 Seconds
(without alignments)
1014.815 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGGQAQMLCIDNCGA.....GRCLSQILHTFEKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 323636 seqs, 103855142 residues

Total number of hits satisfying chosen parameters: 323636

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	22.3	603	5	US-09-573-655B-2154
2	357	22.3	603	5	US-09-573-655B-2192
3	333.5	20.9	282	6	US-10-155-881-9626
4	314	19.6	510	6	US-10-155-881-9875
5	248	15.5	53	6	US-10-108-698-6932
6	232.5	14.5	896	6	US-10-155-881-20031
7	205	12.8	359	6	US-10-155-881-9183
8	165	10.3	143	6	US-10-155-881-9877
9	159	9.9	120	6	US-10-138-145-1127
10	117.5	7.3	200	6	US-10-155-881-9182
11	115	7.2	1774	1	PCT-US02-01339-6
12	94.5	5.9	538	1	PCT-US02-10812-17
13	94.5	5.9	577	1	PCT-US02-10812-15
14	92.5	5.8	298	6	US-10-155-881-33654
15	92.5	5.8	580	6	US-10-155-881-11197
16	92.5	5.8	621	6	US-10-155-881-11230
17	92.5	5.8	753	6	US-10-155-881-25969
18	92.5	5.8	809	6	US-10-155-881-11231
19	92.5	5.8	1015	6	US-10-155-881-28881
20	91.5	5.7	554	6	US-10-155-881-22871
21	91.5	5.7	732	6	US-10-155-881-34231
22	90	5.6	3298	7	US-60-365-384-181
23	90	5.6	3332	7	US-60-365-384-182
24	89	5.6	574	7	US-60-360-039-5730
25	88.5	5.5	260	6	US-10-155-881-10280
26	87.5	5.5	955	6	US-10-154-675-14

ALIGNMENTS

RESULT 1

US-09-573-655B-2154
; Sequence 2154, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2154
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-2154

Query Match 22.3%; Score 357; DB 5; Length 603;
Best Local Similarity 35.2%; Pred. No. 8.6e-24;
Matches 88; Conservative 38; Mismatches 66; Indels 58; Gaps 12;

QY	36	SSFPQSELPKPMTTLVGRLLVPKLNLTQVDNGALPSAVNGA----	AFPSGPAQQGP- 90
DB	60	SSVPSTIRP-GMTIAGQVTVRPTLPMTTMSNPSSQIYNAPIRHPIESP	KARGPR 118
QY	91	PKI-----	TLGYCDCFSGGDFCNSCSC-----NLRHLELRPKAIK 127
DB	119	PNVEGRDGPQPKKQCNKHSRCLKLYCECFASGTGDCGNCVNCFNNDNEP	PARREAVE 178
QY	128	ACLDNRPEAFQPKMGK-----	LGAAKL--RHSGKCNCKRSGCLKNKCYCEYEAKI 176
DB	179	ATERNPFAFRPKRIASSPHGGRRDKREDIGEVLLGKHNGKCHCKSGCL	KYCECFQANI 238
QY	177	MCSSTICKCTACKNYESPSPKRLMSTPH-----	YMB-----PGDFESSHYL-SPAKFS 223
DB	239	LCSENCCLDKCNFEGSEERQALFGEHSNHMAYLQQAANAATGAVGSSG	FAPSPA--- 295
QY	224	GPKLKRKNRQ 233	
DB	296	--PKRRKGQE 303	

RESULT 2

US-09-573-655B-2192
; Sequence 2192, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim

```

; APPLICATION: RUDEN ET AL.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Poly
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106.698

```

ORGANISM: *homo sapiens*

[illegible]

; SEQ ID NO 1127
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-138-145-1127

Query Match 9.9%; Score 159; DB 6; Length 120;
Best Local Similarity 38.7%; Pred. No. 5.9e-07;
Matches 41; Conservative 13; Mismatches 34; Indels 18; Gaps 7;

QY 98 YDCFSGGDFCSCN---NLRHELRFKAKACLDNRNPAFQPKMGKGRIGA-----148
Db 16 YCEC-RAGVYCEPSCQCGCLNKPHEEIVLSTRKQIEFRNPLA-APKVIRLSDAAQETQE 73

QY 149 -----AKLRHSGCKNCRSGCLKNYCEYEAKEIMCSICKCIACKN 189
Db 74 DPNTPASARHKRCNCKKSSCLK-YCEYQGGVGLTNCR-ECNK 117

RESULT 10
US-10-155-881-9182
; Sequence 9182, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9182
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9182

Query Match 7.3%; Score 117.5; DB 6; Length 200;
Best Local Similarity 24.7%; Pred. No. 0.0056;
Matches 39; Conservative 15; Mismatches 57; Indels 47; Gaps 4;

QY 39 PQSELPRKMTTLVG-----RLLPVP--AKNLITQVDNGALPSAVNGAAPSPGALQGGP 91
Db 26 PQAAAPVMDADVAGATNQWPEPDDAHLSPPPPPPPASAPTAAEAAAPTASAAVSP 85

QY 92 KITLSGYCDPSSGDFCNCSNNLRHELRFKAKACLDNRNPAFQPKMGKGRIGA 151
Db 86 -----APAEANGNSDRKRRKADDDGG-- 107

QY 152 RHSGCKNCRSGCLKNYCEYEAKEIMCSICKCIACKN 189
Db 108 --CRACCKSRCLKLYCVYASGSHCTELCGCEPCHN 143

RESULT 11
PCT-US02-01339-6
; Sequence 6, Application PC/TUS0201339
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: WALIA, Narinder K.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom

; APPLICANT: YUE, Henry
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra
; APPLICANT: WARREN, Bridget A.
; APPLICANT: YAO, Monique G.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: LU, Yan
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Daniel B.
; APPLICANT: GRAUL, Richard
; APPLICANT: LU, Dyung Aina M.
; TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
; FILE REFERENCE: PI-0346 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/01339
; CURRENT FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/262,838; 60/265,927; 60/271,196; 60/274,549; 60/334,1
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 1774
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7722591CDI
PCT-US02-01339-6

Query Match 7.2%; Score 115; DB 1; Length 1774;
Best Local Similarity 23.1%; Pred. No. 0.13;
Matches 73; Conservative 27; Mismatches 124; Indels 92; Gaps 18;

QY 17 NCGARELKALHLLPQYDDQSSFFQSELPKPMPTLVGRLLPVPKAKNLITQVDNGALPSAV 76
Db 748 NCNA-----WLLPDLTRSAS-----VGPPEERSVAHAAVAGSRLYISGGFGGVALGRL 796

QY 77 NGAAPSGPA--LQGPPIKTLGYCD-CFS---SGD-----FCNSCSCNNL---RHELER 122
Db 797 LALTLPPDPCRLILSSPEACNOSGACTWCHGACLSGDDAHLGCGGSPCPMPSPSPCECRR 856

QY 123 FKAIKACLDNRNPAFQPKMGKGRIGAALRHSKGC-NCKRSGCL-----KNYC----- 169
Db 857 LRTCSECLARHPTLQPDGE-----ASTPRCKWCTNCPGACIGRNGSCTSENDCRINQ 911

QY 170 -----ECYEAKI-----MCSICKIACKNYEESPE-RKMLMSTPHY----- 205
Db 912 REVFWAGNCSEAAAGAACDEQCTREGKCMWTRQPKRTGTRRILSVQPTYDWTCTFSHLL 971

QY 206 -MEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISWEWEATCATCLLAOGEEAEQHCSP 264
Db 972 NVSPMPVESS---PPLPCPTPCHELLPN-----CYS-----CLDSKGADGGQHCWV 1014

QY 265 SLAEOMILEEFGRCLS 280
Db 1015 SSSLQ-----QCLS 1023

RESULT 12
PCT-US02-10812-17
; Sequence 17, Application PC/TUS0210812
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: DELEGEANE, Angelo M.

APPLICANT: SWARNAKAR, Anita
APPLICANT: HAFALIA, April J.A.
APPLICANT: DUGGAN, Brendan M.
APPLICANT: WARREN, Bridget A.
APPLICANT: EMERLING, Brooke M.
APPLICANT: ARVIZU, Chandra S.
APPLICANT: HONCHELL, Cynthia D.
APPLICANT: KALLICK, Deborah A.
APPLICANT: LU, Dyung Aina M.
APPLICANT: LEE, Ernestine A.
APPLICANT: YUE, Henry
APPLICANT: FORSYTHE, Ian J.
APPLICANT: RAMKUMAR, Javalaxmi
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: LI, Joana X.
APPLICANT: THANGAVELU, Kavitha
APPLICANT: BAUGHN, Mariah R.
APPLICANT: YAO, Monique G.
APPLICANT: SANJANWALA, Madhu M.
APPLICANT: WALIA, Narinder K.
APPLICANT: BURFORD, Neil
APPLICANT: LAL, Preeti, G.
APPLICANT: BECHA, Shanya D.
APPLICANT: LEE, Soo Y.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: LUO, Wen
APPLICANT: LU, Yan
APPLICANT: WANG, Yu-mei E.
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PI-0397 PCT
CURRENT APPLICATION NUMBER: PCT/US02/10812
CURRENT FILING DATE: 2002-04-08
PRIOR APPLICATION NUMBER: 60/282,282; 60/283,782; 60/284,823; 60/287,264;
60/288,662; 60/290,383; 60/298,348; 60/351,928;
60/359,903
PRIOR FILING DATE: 2001-04-05; 2001-04-13; 2001-04-18; 2001-04-27;
2001-05-04; 2001-05-11; 2001-06-15; 2002-01-25;
2002-02-25
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PERL Program
SEQ ID NO 17
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7506402CD1
PCT-US02-10812-17

Query Match 5.9%; Score 94.5; DB 1; Length 538;
Best Local Similarity 20.5%; Pred. No. 2.1;
Matches 62; Conservative 37; Mismatches 110; Indels 93; Gaps 17;

QY 14 CIDNCGARELKAHLHPQYDDOSSPQSELPKPMTTLVGRLLPVPAKLNLTQVDNGALP 73
DB 76 CYSNCSPPKFQILML-----FPNLYPKET-----LEAFVIVTQM----- 112

QY 74 SAVNGAAPPSPGALQGPCKITLSGYDCFSFGDFCN-----SCSCNNLRHEL 120
DB 113 -----LALSLSGISYDDPKK-----CQSESTCINNPEVVSQNGVKTFSSCSLSRSFQFI 161

QY 121 ERFAKAKCLDRNPAFQPKMGKGRGLRAKLRHSGCKCK-----RSGLKN 167
DB 162 SNV-GVK-CLQNKPKQ-MOKSKPKPCVGNRLEGNICDGTAEACQGPASCDFRTCVLKD 218

QY 168 YCECYEAKIMCSICKC-----LACKNYEESPERKMLMSTPPHYMEPGDRESSHYLSPAKFS 223
DB 219 GAKCY--KGLCKKDCQILQSGVECRP-KAHPECDTA-----ENCNGSSPE--C 261

QY 224 GPP-----KLKNRQAFSCISNEV--VEATCACLLAQGEAEARQHCSPSLAEQMIIEFG 276
DB 262 GPDITLNLGCKNKNKFCYDGDCHDLDAKCESVFGKSGSRNAPFACYEIQSQS--DREG 319

QY 277 RC 278
DB 320 NC 321

RESULT 13

PCT-US02-10812-15
; Sequence 15, Application PC/TUS0210812
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: GANDHI, Ameen R.
; APPLICANT: DELEGEANE, Angelo M.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: YUE, Henry
; APPLICANT: FORSYTHE, Ian J.
; APPLICANT: RAMKUMAR, Javalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: LI, Joana X.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YAO, Monique G.
; APPLICANT: SANJANWALA, Madhu M.
; APPLICANT: WALIA, Narinder K.
; APPLICANT: BURFORD, Neil
; APPLICANT: LAL, Preeti, G.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: LEE, Soo Y.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: LUO, Wen
; APPLICANT: LU, Yan
; APPLICANT: WANG, Yu-mei E.
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PI-0397 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/10812
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/282,282; 60/283,782; 60/284,823; 60/287,264;
60/288,662; 60/290,383; 60/298,348; 60/351,928;
60/359,903
; PRIOR FILING DATE: 2001-04-05; 2001-04-13; 2001-04-18; 2001-04-27;
2001-05-04; 2001-05-11; 2001-06-15; 2002-01-25;
2002-02-25
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 55064352CD1
PCT-US02-10812-15

Query Match 5.9%; Score 94.5; DB 1; Length 577;
Best Local Similarity 20.5%; Pred. No. 2.3;
Matches 62; Conservative 37; Mismatches 110; Indels 93; Gaps 17;

QY 14 CIDNCGARELKAHLHPQYDDOSSPQSELPKPMTTLVGRLLPVPAKLNLTQVDNGALP 73
DB 76 CYSNCSPPKFQILML-----FPNLYPKET-----LEAFVIVTQM----- 112
QY 74 SAVNGAAPPSPGALQGPCKITLSGYDCFSFGDFCN-----SCSCNNLRHEL 120

Db 113 -----LALSIGISYDDPKK-----COCSESTCMMPEVVQSNVKTFFSSCSLSRSTQFI 161
QY 121 ERFKAIAKACLDNRNPEAFQPKMGKRLGAALRHSGKCNCK-----RSGCLKN 167
Db 162 SNV-GVK-CLQNKPO-WQKSPKPCVCGNGLENEICDGTAEACGPGASCCDPRTCVLKD 218
QY 168 YCEYAKIMCSICKC-----IACKNYEESPERRKMLMSTPHYMEPGDFESSHYLSPAKFS 223
Db 219 GAKCY--KGLCCCKDCQILQSGVECRP-KAHPECDIA-----ENCNGSSPE--C 261
QY 224 GPP-----KLKRNQAFSCISWEV--VEATCACLQAQEEAEQEHGSPSLAEQMIILEFG 276
Db 262 GPDTITLGLSKNNKFCYDGDCHLDARCESVFGKGRNAPFACYEEIQSOS--DRFG 319
QY 277 RC 278
Db 320 NC 321

RESULT 14
US-10-155-881-33654
; Sequence 33654, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: TRANSCRIPTION IN PLANTS
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 33654
; LENGTH: 298
; TYPE: PRT
; ORGANISM: zea mays
US-10-155-881-33654

Query Match 5.8%; Score 92.5; DB 6; Length 298;
Best Local Similarity 23.7%; Pred. No. 1.6; Mismatches 29; Indels 49; Gaps 7;
Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;
QY 94 TLSGYDCFSFGDFCNS-CSCNNLRHELERFKAIAKACLDNRNPEAFQPKMGKRLGAALR 152
Db 53 TCGKQCPCLSNGTCCKEYCG-----PKICKNRF----- 81
QY 153 HSKGCNCKRSGCLKNYCEYEAkimcs-SICK-C-IAC-----KNYE 191
Db 82 --RGCHCAKSQCRSROCPFAADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 133

RESULT 15
US-10-155-881-11197
; Sequence 11197, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: TRANSCRIPTION IN PLANTS
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 11197
; LENGTH: 580
; TYPE: PRT
; ORGANISM: zea mays

US-10-155-881-11197
Query Match 5.8%; Score 92.5; DB 6; Length 580;
Best Local Similarity 23.7%; Pred. No. 3.5;
Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;
QY 94 TLSGYDCFSFGDFCNS-CSCNNLRHELERFKAIAKACLDNRNPEAFQPKMGKRLGAALR 152
Db 335 TCGKQCPCLSNGTCCKEYCG-----PKICKNRF----- 363
QY 153 HSKGCNCKRSGCLKNYCEYEAkimcs-SICK-C-IAC-----KNYE 191
Db 364 --RGCHCAKSQCRSROCPFAADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 415
Search completed: July 11, 2002, 08:14:15
Job time: 39 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:16:59 ; Search time 53.06 Seconds
(without alignments)
625.915 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQLIHTFEKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802:*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	100.0	299	21	AA198464 Human testis speci
2	1251	76.0	295	21	AA198463 Mouse testis speci
3	470	28.5	950	22	ABB62035 Drosophila melanog
4	446	27.1	147	22	AA195330 Human reproductive
5	441	26.8	438	20	AA198392 Caenorhabditis ele
6	378	23.0	280	22	ABG17958 Novel human diagno
7	278	16.9	53	22	AA198158 Human colon cancer
8	216	13.1	251	22	AA198348 Human polypeptide,
9	200	12.1	243	22	ABB68888 Drosophila melanog
10	113	6.9	1981	22	ABB61657 Drosophila melanog
11	109.5	6.6	699	20	AA1981840 Human LFA-1 beta c

12	109.5	6.6	769	9	AA1980836 Beta subunit of hu
13	109.5	6.6	769	11	AA19807113 Recombinant beta-s
14	109.5	6.6	769	16	AA1980108 LFA-1 CD18 polyep
15	109.5	6.6	769	21	AA19803974 LFA-1 CD18 polyep
16	109.5	6.6	793	22	AB196399 Novel human diagno
17	108.5	6.6	769	13	AA19824256 Beta-subunit CD18
18	108	6.6	676	18	AA19835853 Human CD18 for use
19	108	6.6	3011	13	AA19821519 Compiled HCV seque
20	105.5	6.4	695	22	ABB30979 Peptide #3630 enco
21	105.5	6.4	695	22	ABB36170 Peptide #3676 enco
22	105.5	6.4	695	22	AA19856953 Human brain expres
23	105.5	6.4	695	22	AA19869338 Human bone marrow
24	105.5	6.4	695	22	AA19817170 Peptide #3604 enco
25	105.5	6.4	695	22	AA19829661 Peptide #3698 enco
26	105.5	6.4	695	22	AA19804863 Peptide #3545 enco
27	104	6.3	3011	14	AA19831621 Hepatitis C virus
28	103.5	6.3	1725	21	AA19819800 Mouse laminin 2 ma
29	103.5	6.3	1725	21	AA19848451 Mouse laminin 8 po
30	103.5	6.3	1764	10	AA1981672 Primary amino acid
31	103.5	6.3	1776	19	AA19850894 Mouse laminin B1 c
32	103.5	6.3	1786	21	AA19819799 Mouse laminin 2 be
33	103.5	6.3	1786	21	AA19848450 Mouse laminin 8 po
34	102	6.2	1788	22	ABB62995 Drosophila melanog
35	101.5	6.2	762	22	ABG08032 Novel human diagno
36	101	6.1	507	16	AA19866631 HCV J1 NS3-NS4 dom
37	101	6.1	609	15	AA19851170 Hepatitis C virus
38	101	6.1	631	18	AA19831884 A nonstructural pr
39	101	6.1	686	22	AA19862633 HCV NS3A polyep
40	101	6.1	1021	18	AA19834481 HCV antigen combin
41	101	6.1	1021	19	AA19840039 Fusion protein c20
42	101	6.1	1771	22	AA19862631 HCV NS35 polyep
43	101	6.1	1771	22	AA19862632 HCV delNS35 polyep
44	101	6.1	1771	22	AA19862634 Amino acid sequenc
45	101	6.1	1771	22	AA19862635 Amino acid sequenc

ALIGNMENTS

RESULT 1

AA198464

ID AA198464 standard; Protein; 299 AA.

XX

AC AA198464;

XX

DT 25-APR-2000 (first entry)

XX

DE Human testis specific factor tesmin SEQ ID NO:5.

XX

Testis specific factor; tesmin; cell death; regulation; spermatocyte;

KW differentiation regulatory factor; male germ cell regulatory actor;

KW germ cell differentiation; sterility.

XX

OS Homo sapiens.

XX

PN WO200004147-A1.

XX

PD 27-JAN-2000.

XX

PF 16-JUL-1999; 99WO-JP03859.

XX

PR 17-JUL-1998; 98JP-0219856.

XX

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX

PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX

DR WPI: 2000-147785/13.

DR N-PSDB; AA1986156.

XX

PT New male germ cell regulatory factor tesmin expressed in spermatocytes useful for investigation of germ cell differentiation and sterility -

XX PS Claim 1; Page 50-52; 63pp; Japanese.

XX CC The present sequence represents a male germ cell regulatory factor

XX CC expressed specifically in spermatocytes, designated tesmin.

XX CC CC can be used in the investigation of the mechanisms of germ cell

XX CC differentiation and sterility.

XX SQ Sequence 299 AA;

Query Match 100.0%; Score 1647; DB 21; Length 299;

Best Local Similarity 100.0%; Pred. No. 2.4e-141;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVICQLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYLOSDVPKPMALVGRFLPASTK 60

Db 1 mviqqlkggtqmlcidsrtrelkalhlvpqyqdnnylqsdvpkpmalvgrflpastk 60

Qy 61 LNLITQLEGALPSVWNGSAFPGSTLPGPKITLAGYCDCFASGDFCNCCNCCNNL 120

Db 61 lnlitqlegalpsvngsafpgstlpgpkitlagycdcfsgdfcnccnccnnl 120

Qy 121 HHDIERFAIKACLRNPEAFQPKIGKQLGNVQHNKGCNCRSSGCLKNYCEYEAQI 180

Db 121 hhdiierfaiakclrnpeafqpkigkqlgnvqhnkgcncrrsgclknyceyeaqi 180

Qy 181 MCSSICKICGCKNYEESPERKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240

Db 181 mcssickicgcknyeesperktlmsmpnmydmotglegshylypptkfsglprfshdrpss 240

Qy 241 CISWEVVEATCACLLOAEAEKHCSCLAEQMILEFGRCLSQIILHTEFKSKGLKME 299

Db 241 ciswevveatcaclllaeaeekhcscklaeqmileeefgrclsqilhteifksgklkme 299

RESULT 2

AY68463

ID AAY68463 standard; Protein; 295 AA.

AC AAY68463;

DT 25-APR-2000 (first entry)

DE Mouse testis specific factor tesmin SEQ ID NO:4.

KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;

KW differentiation regulatory factor; male germ cell regulatory actor;

KW germ cell differentiation; sterility.

OS Mus musculus.

PN WO200004147-A1.

XX 27-JAN-2000.

XX 16-JUL-1999; 99WO-JP03859.

XX 17-JUL-1998; 98JP-0219856.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX WPI; 2000-147785/13.

DR N-PSDB; AAZ88155, AAZ88157.

XX New male germ cell regulatory factor tesmin expressed in spermatocytes

PT useful for investigation of germ cell differentiation and sterility -

XX PS Claim 1; Page 47-49; 63pp; Japanese.

XX CC The present sequence represents a male germ cell regulatory factor

XX CC expressed specifically in spermatocytes, designated tesmin.

XX CC CC can be used in the investigation of the mechanisms of germ cell

XX CC differentiation and sterility.

XX SQ Sequence 295 AA;

Query Match 76.0%; Score 1251; DB 21; Length 295;

Best Local Similarity 76.3%; Pred. No. 2.1e-105;

Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICQLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYLOSDVPKPMALVGRFLPASTK 60

Db 1 mviqqlkggtqmlcidsrtrelkalhlvpqyqdnnylqsdvpkpmalvgrflpastk 60

Qy 61 LNLITQLEGALPSVWNGSAFPGSTLPGPKITLAGYCDCFASGDFCNCCNCCNNL 120

Db 61 lnlitqlegalpsvngsafpgstlpgpkitlagycdcfsgdfcnccnccnnl 116

Qy 121 HHDIERFAIKACLRNPEAFQPKIGKQLGNVQHNKGCNCRSSGCLKNYCEYEAQI 180

Db 117 rhelerfaiakclrnpeafqpkigkqlgnvqhnkgcncrrsgclknyceyeaki 176

Qy 181 MCSSICKICGCKNYEESPERKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240

Db 177 mcssickicgcknyeesperktlmsmpnmydmotglegshylypptkfsglprfshdrpss 236

Qy 241 CISWEVVEATCACLLOAEAEKHCSCLAEQMILEFGRCLSQIILHTEFKSKGLKME 299

Db 237 ciswevveatcaclllaeaeekhcscklaeqmileeefgrclsqilhteifksgklkie 295

RESULT 3

ABB62035

ID ABB62035 standard; Protein; 950 AA.

XX ABB62035;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 12897.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL06138.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-465570/50.
XX N-PSDB; AAL01300.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen -
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.
XX
XX Sequence 147 AA;
SQ

Query Match 27.1%; Score 446; DB 22; Length 147;
Best Local Similarity 98.9%; Pred. No. 1.1e-32;
Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MVTICQLKGGTQMLCINSTRRELKALHLVPOYQDQNNYLOSDVPKPMFTALVGRFLPASTK 60
Db 60 mvicqlkggtqmlcidsnrtrelkalhlvpqyqdgqnylqsdvvpkpmftalvgrflpastk 119
QY 61 LNLITQLEGALPSVVNGSAFPGSTLP 88
Db 120 Inltqlegalpsvvngsafpgstlp 147
RESULT 5
AAW83392
ID AAW83392 standard; Protein; 438 AA.
XX
XX AAW83392;
AC
XX 29-MAR-1999 (first entry)
DT
XX
XX Caenorhabditis elegans synMuv protein LIN-54.
DE
XX
XX LIN-54; synthetic multivulvar; SynMuv; signal transduction;
KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
KW cell proliferation; gene therapy.
XX
XX Caenorhabditis elegans.
XX
XX WO9854299-A1.
XX
XX 03-DEC-1998.
XX
XX 28-MAY-1998; 98WO-US11043.
XX
XX 28-MAY-1997; 97US-0047996.
XX
XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Ceol C, Horvitz HR, Lu X;
XX
XX WPI: 1999-045362/04.
XX N-PSDB; AAV72865.
XX
XX Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
PT useful for treating diseases associated with altered levels of cell
XX proliferation, e.g. carcinomas
XX
XX Claim 7; Fig 13; 70pp; English.
XX
XX This is the amino acid sequence of LIN-54, a novel protein of
CC Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
CC synthetic multivulvar (synMuv) gene involved in cell fate and cell
CC proliferation, and is part of a pathway that may be used as a
CC genetic and biochemical model system for tumour suppression and
CC cancer in mammals. SynMuv pathway genes and proteins may be used
CC to identify genes which are part of the mammalian pathway and
CC to modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
CC C. elegans SynMuv polypeptides selected from LIN-37, -35, -55, -52,
CC -53, -54 and E2f-1 (see AAW83386-92), are new. Also claimed are:
CC (1) nucleic acids (see AAV72866-67) encoding the mouse and human
CC homologues of C. elegans LIN-54; (2) vectors containing the nucleic
CC acids; (3) transgenic cells; (4) a pure mammalian SynMuv
CC polypeptide; and (5) an antibody which binds to a SynMuv family
CC protein. The SynMuv nucleic acids and polypeptides can be used to
CC diagnose and treat, especially by gene therapy, conditions
CC involving altered levels of cell proliferation, e.g. SynMuv-
CC associated carcinomas.
XX
XX Sequence 438 AA;
SQ

CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 53 AA;

Query Match 16.9%; Score 278; DB 22; Length 53;
Best Local Similarity 100.0%; Pred. No. 5.5e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 VEATCACLAAQGEAEKEHCKSLAQOMILEEFGRCLSQILHTFEFKSKGLKME 299
|||||
DB 1 veatcaccaggaeeakehckslaeqmileefgrclsgilhtefksgikme 53

RESULT 8

AAW93348
ID AAW93348 standard; Protein: 251 AA.

XX
AC AAW93348;

XX
DT 06-NOV-2001 (first entry)

XX
DE Human polypeptide, SEQ ID NO: 2895.

XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.

XX
OS Homo sapiens.

XX
PN EP130094-A2.

XX
PD 05-SEP-2001.

XX
PF 07-JUL-2000; 2000EP-0114089.

XX
PR 08-JUL-1999; 99JP-0194486.

XX
PR 11-JAN-2000; 2000JP-0118774.

XX
PR 02-MAY-2000; 2000JP-0183765.

XX
PA (HELI-) HELIX RES INST.

XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX
DR WPI: 2001-524255/58.

XX
DR N-PSDB; AAK94268.

XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX
PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.

XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide

CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 251 AA;

Query Match 13.1%; Score 216; DB 22; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.6e-11;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTCQLKGGTQMCLIDNSRTRELKALHLVPPQYODQNNYLOS 41
|||||
DB 210 mvicqlkkggtqmclidnsrtrelkalhlvppqyodqnnylqs 250

RESULT 9

ABB68888

ID ABB68888 standard; Protein: 243 AA.

XX
AC ABB68888;

DT 26-MAR-2002 (first entry)

XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33456.

XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX
OS Drosophila melanogaster.

XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.

XX
PF 23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.

XX
PR 11-JUL-2000; 2000US-0614150.

XX
PA (PEKE) PE CORP NY.

XX
PI Venter JC, Adams M, Li PWD, Myers EW;

XX
DR WPI: 2001-656860/75.

XX
DR N-PSDB; ABL12991.

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX
PS Disclosure; SEQ ID NO 33456; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBV2072).

XX
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 243 AA;

Query Match 12.1%; Score 200; DB 22; Length 243;
Best Local Similarity 25.5%; Pred. No. 4.4e-10;
Matches 59; Conservative 25; Mismatches 63; Indels 84; Gaps 7;

QY 136 RNPEAFQPKIGKGO-LGNVKPQHNKGNCRRRSGCLKNYCEYEAQIMCSSICKIGCKNY 194

Db 7 rsvdkadgkkgaggv-----kccckrqckiknydcygsmaictkfcrcvgrnt 61
QY 195 E-----ESPER-----KTL 203
Db 62 evrelvdpsvaksavkrqkaamsakaaakagidvgkalqvaastlalpgkal 121
QY 204 MSMPNYMTGG---LEGSHYLP-----PTKFSGLPRFS-----HDR 236
Db 122 mppkytlvagkppmasshinpipsrpiataatparavkqaeppmpvnliipvrhddr 181
QY 237 RPSSCISWEVATACCLLAQGEAEKHCCKLAEQMILEFGRCLSOIL 287
Db 182 rdrrnifvqpvnaalliecmilgateaeqlglnelqvclvleefmrgyknll 232

RESULT 10
ABB61657
ID ABB61657 standard; Protein; 1981 AA.
XX
AC ABB61657;
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11763.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO2001171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEXE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL05760.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 11763; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1981 AA;

Query Match 6.9%; Score 113; DB 22; Length 1981;
Best Local Similarity 21.3%; Pred. No. 0.47; Indels 58; Gaps 7;
Matches 44; Conservative 26; Mismatches 79; Indels 58; Gaps 7;

QY 44 PKPMTALVGRFLPASTKLNLTQLEALPSVVGSAFPSPGSLPGPPKITLAGYDCDCA 103

Db 356 phttsilkytpss---nsntrnsnmpstntssstetdsal----- 396
QY 104 SGDFCNMCN-----NCCNNLHHDIERFAKACLCGRNPEAFQPKIGKQLGNVYPOHN 158
Db 397 -----cscnptrfntldfrsnannrtytklnlclsnssqsis---snsqcsnkslcn 446
QY 159 KGCNRRSCLKNYCYEAQIMC-----SSICKICGCKNYEES----- 197
Db 447 stcnssnslcnsssssklcsnsgysssksklnsgysssksklnpsnpslyfa 506
QY 198 PERKTLMSMP---NYMTGGLEGSHYL 221
Db 507 pgtklliesivaeinvnktgtgfenshkl 533

RESULT 11
AAW81840
ID AAW81840 standard; Protein; 699 AA.
XX
AC AAW81840;
DT 05-FEB-1999 (first entry)
XX
DE Human LFA-1 beta chain protein.
XX
KW LFA-1; beta chain; human; anti-inflammatory; soluble.
XX
OS Homo sapiens.
XX
PN JP10295375-A.
PD 10-NOV-1998.
XX
PF 25-APR-1997; 97JP-0109752.
XX
PR 25-APR-1997; 97JP-0109752.
XX
PA (DAUC) DAIICHI PHARM CO LTD.
XX
XX WPI; 1999-038275/04.
DR N-PSDB; AAV64856.
XX
PT Nucleic acid encoding soluble LFA-1 protein - used to identify
PT anti-inflammatory agents for use with LFA-1 mediated activity
XX
PS Disclosure; Page 13; 14pp; Japanese.
XX
CC This sequence represents a soluble LFA-1 beta chain protein. The
CC encoding nucleic acids can be used for the evaluation of an
CC anti-inflammatory agent by a test for the combination between
CC ICAM protein and a soluble LFA-1 hetero dimer protein. The soluble
CC LFA-1 protein does not associate in the absence of a surfactant.
XX
SQ Sequence 699 AA;

Query Match 6.6%; Score 109.5; DB 20; Length 699;
Best Local Similarity 22.3%; Pred. No. 0.27;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 WICQLKGGTQMLCINDSRTRELKALHLPQYQDQNNYLQSDVPKPMALVGRFLPASTK 60
Db 436 itvtvlpqcecrdrqsrsl--ch-----gkfglecicrddtyigkncecqtq 486
QY 61 LNLITQOLEGA-----LPSVVGSAFPSPGSLPGPPKITLAGYDCDCA----- 103
Db 487 -grssgelegscrkdnslsicsglgdcvcgqlchsdvpg--kliygycecdtincer 543
QY 104 -SGDFCN-----CNCNCCNNLHHDIERF-----KAKACLCGRNPEAFQPKIGKQLGN 152
Db 544 yngqvccggpgrglcfcgkc--rchpgefsgacqcertegcl--nprrrvecs-grgr--- 595

Qy 153 VKPOHKGCCNRRSGCLKNYCEVYEAQIM-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLCQCPGPCSPGKYISCAECLKFEKGP 638

RESULT 12
 AAP80836
 ID AAP80836 standard; protein; 769 AA.
 AC AAP80836;
 XX
 DT 18-OCT-1990 (first entry)
 XX
 DE Beta subunit of human glycoproteins LFA-1, Mac-1 or p150,95
 DE encoded by cDNA clone 18.1.1.
 XX
 KW Immuno-assays; autoimmune diseases; allograft rejection;
 KW graft-versus-host disease; leukocyte adhesion deficiency; cDNA clone;
 KW beta-subunit of human glycoproteins LFA-1, Mac-1 or p150,95.
 XX
 OS Homo sapiens.

PH Key Location/Qualifiers
 FT Modified-site 50
 FT Modified-site /note="Potential N-glycosylation site"
 FT Modified-site 116
 FT Modified-site /note="As above"
 FT Modified-site 212
 FT Modified-site /note="As above"
 FT Modified-site 244
 FT Modified-site /note="As above"
 FT Modified-site 501
 FT Modified-site /note="As above"
 FT Modified-site 642
 FT Modified-site /note="As above"
 FT Region 197..205
 FT Region /label=TRYPTIC PEPTIDE L-56a
 FT Region 258..269
 FT Region /label=TRYPTIC PEPTIDE M-58
 FT Region 311..324
 FT Region /label=TRYPTIC PEPTIDE P-61
 FT Region 340..351
 FT Region /label=TRYPTIC PEPTIDE M-52
 FT Region 367..375
 FT Region /label=TRYPTIC PEPTIDE L-65
 FT Region 472..479
 FT Region /label=TRYPTIC PEPTIDE P-20
 FT Region 489..498
 FT Region /label=TRYPTIC PEPTIDE P-18
 FT Region 528..539
 FT Region /label=TRYPTIC PEPTIDE L-56b

PN W08806592-A.
 XX
 XX 07-SEP-1988.
 XX
 XX 26-FEB-1988; 88WO-US00611.
 XX
 XX 26-FEB-1987; 87US-0019440.
 XX
 XX (DANA-) DANA FARBER CANCER.
 XX
 XX Springer T, Kishimoto T, Roberts T;
 XX
 XX WPI: 1988-271137/38.
 XX
 XX N-PSDB: AAN80863.
 XX
 XX Recombinant beta-sub-unit of human glycoprotein -
 XX used for treatment and diagnosis of diseases involving cellular
 XX adhesion and for producing antibodies
 XX
 XX ; Fig 1; 22pp; English.
 XX

CC The beta-subunit of the glycoproteins was isolated and the AA sequence of
 CC at least a portion of it detd. A synthetic oligo probe corresp. to the AA
 CC sequence was synthesised and used as probe for a genomic or cDNA library
 CC contg. a DNA sequence encoding the beta-subunit. The p150,95 protein was
 CC purified from hairy cell leukaemia spleens. Mac-1 was purified from
 CC leukocytes and LFA-1 was purified from the SKW3 T cell line. The cDNA can
 CC be used to produce recombinant beta-subunit in large amts. Monoclonal
 CC antibodies raised against the recombinant beta-subunits can be used in
 CC immunoassays to monitor diseases such as AIDS, autoimmune disease,
 CC allograft rejection, graft-versus-host disease and leukocyte adhesion
 CC deficiency (LAD). The beta-subunit cDNA clone can be used in prenatal
 CC diagnosis of LAD. Diagnosis of LAD uses a 1.8Kb EcoRI fragment encoding
 XX the beta-subunit of human LFA-1 as a probe.
 XX

Sequence 769 AA;

Query Match 6.6%; Score 109.5; DB 9; Length 769;
 Best Local Similarity 22.3%; Pred. No. 0.3;
 Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

Qy 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYVLOSVDPKPMTALVGRFLPASTK 60
 Db 436 ivtvqlpqecrcrdqsrdrsl--ch-----gkfglegicrdctgyigkncecqtq 486
 Qy 61 LNLITQOLEGA-----LPSVVNGSAFSPGSTLPGPPKITLAGYDCDFA----- 103
 Db 487 -grsselegscrkdnnsilcsglgdcvcqclchtsdvpg--klyggycecdtincer 543
 Qy 104 -SGDFCNN-----CNCNNCCNNLHDIETRF-----KAIKACLGRRNPEAFQPKIGKGQLGN 152
 Db 544 yngqvcgpggrglcfcgkc--rchpgfegsacqcerttegcl--nprrvces-grgr--- 595
 Qy 153 VKPOHKGCCNRRSGCLKNYCEVYEAQIM-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLCQCPGPCSPGKYISCAECLKFEKGP 638

RESULT 13

AAR07113
 ID AAR07113 standard; protein; 769 AA.

XX AAR07113;

XX 24-JAN-1991 (first entry)

XX Recombinant beta-subunit of LFA-1.

XX Rhinovirus; LFA-1; Mac-1; p150,95; ICAM-1; autoimmune diseases;
 XX graft rejection.

Key	Location/Qualifiers
Peptide	197..205
Peptide	/label=tryptic peptide L-56a
Peptide	258..269
Peptide	/label=tryptic peptide H58
Peptide	311..324
Peptide	/label=tryptic peptide P-61
Peptide	340..351
Peptide	/label=tryptic peptide H-52
Peptide	367..375
Peptide	/label=tryptic peptide L-65
Peptide	472..479
Peptide	/label=tryptic peptide P-20
Peptide	489..498
Peptide	/label=tryptic peptide P-18
Peptide	528..539
Peptide	/label=tryptic peptide L-56b

PN W09010652-A.

XX 20-SEP-1990.

XX

Fri Jul 12 08:35:30 2002

KW Leukocyte function associated antigen; LFA-1; I domain; inhibitor;
KW ligand; binding; pathology; treatment; screening; detection;
KW inflammatory disease; autoimmune disease; reperfusion injury;
KW myocardial infarction; stroke; organ transplant;
KW respiratory distress syndrome; glomerulonephritis; Crohn's disease;
KW T cell mediated disease; rheumatoid arthritis; osteoarthritis;
KW spondylitis; thyroid associated ophthalmopathy; Behcet disease;
KW sepsis; asthma; chronic bronchitis; silicosis; pulmonary sarcoidosis;
KW fibrosis; cystic fibrosis; keloid formation; scar formation;
KW atherosclerosis; transplant rejection; lupus;
KW inflammatory bowel disease; ulcerative colitis; leukaemia;
KW atopic dermatitis; psoriasis; urticaria; uveitis.

XX Homo sapiens.

XX OS

XX WO200060355-A2.

XX 12-OCT-2000.

XX 03-APR-2000; 2000WO-US08841.

XX 02-APR-1999; 99US-0285477.

XX (ICOS-) ICOS CORP.

XX (ABBO) ABBOTT LAB.

XX Staunton D, Van Der Vieren M, Huth J, Fowler K, Orme M;

XX Olejniczak ET;

XX WPI; 2000-665028/64.

XX N-PSDB; AAA54361.

XX Identifying inhibitor of leukocyte function-associated antigen binding

XX to natural ligand for ameliorating inflammatory diseases, by measuring
XX antigen-ligand binding in presence and absence of test compound
XX
XX Example 4; Page 63-65; 66pp; English.
XX
XX Inhibitors of leukocyte function-associated antigen binding are
XX useful for manufacturing a medicament for ameliorating a pathology
XX arising from LFA-1 binding to its ligand such as inflammatory
XX diseases, autoimmune diseases, reperfusion injury, myocardial
XX infarction, stroke, organ transplant, adult respiratory distress
XX syndrome, acute glomerulonephritis, Crohn's disease, T cell mediated
XX diseases, rheumatoid arthritis, osteoarthritis, spondylitis, thyroid
XX associated ophthalmopathy, Behcet disease, sepsis, asthma, chronic
XX bronchitis, allergic respiratory distress syndrome, chronic pulmonary
XX inflammatory disease, such as chronic obstructive pulmonary disease,
XX silicosis, pulmonary sarcoidosis, fibrosis, cystic fibrosis, keloid
XX formation, scar formation, atherosclerosis, transplant rejection
XX disorders, such as graft versus host reaction and allograft
XX rejection, chronic glomerulonephritis, lupus, inflammatory bowel
XX disease, such as ulcerative colitis, proliferative lymphocyte
XX diseases, such as leukemia, and inflammatory dermatoses, such as
XX atopic dermatitis, psoriasis, urticaria, and uveitis.

XX Sequence 769 AA;

Query Match 6.6%; Score 109.5; DB 21; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.3;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICQLKGTOMLCIDNSRTELKALHLVPOYQDNVNYLQSDVPKPMATLVGRELPASTK 60
DB 436 ivtvqlpqccrctrdqdrsl--ch-----gkfglegicrctdygkncectq 486
QY 61 LNLITQQLLEGA-----LPSVVNGSAFSGSTLPGPPIKITLAGYCDCEFA----- 103
DB 487 -grsqlegrckdnnsiicsglgdcvcgqclchtsdvp--kliyqycccdtincer 543
QY 104 -SGDFCNN-----CNCNCCNNLHHDIERF-----KAIKACIGRNPEAFQPIKIGKOLGN 152

Db 544 yngqvcgpggrglcfcgkc--rchpgfegsqcqcrttegcl--nprrrvecs-grgr--- 595
QY 153 VKPQHNGKGNCRRRSGLLKNYCEYEAQIM-----CSSIC-----KCIGCKNVEESP 198
Db 596 -----crc-----nvcechsgyqlplcqcpgpcpgkyiscaecikfexgp 638

Search completed: July 11, 2002, 08:17:01
Job time: 205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:15:20 ; Search time 54.3 seconds
(without alignments)
529.111 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQILHTEFKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	437	26.5	429	2	T23152	hypothetical prote
2	376	22.8	571	2	B84585	hypothetical prote
3	357	21.7	603	2	T08955	hypothetical prote
4	245.5	14.9	658	2	F71410	hypothetical prote
5	117	7.1	497	2	T27827	hypothetical prote
6	112	6.8	1737	2	T00209	MEGF8 protein - hu
7	109.5	6.6	769	1	LJHULM	leukocyte adhesion
8	106.5	6.5	3635	2	T10053	laminin alpha 5 ch
9	105.5	6.4	2180	2	T29764	hypothetical prote
10	104	6.3	775	2	T21436	hypothetical prote
11	103.5	6.3	1786	1	MMMSB1	laminin beta-1 cha
12	102	6.2	1790	1	MMFEB1	laminin beta-1 cha
13	101	6.1	1700	2	S08167	Balbani ring 3 pr
14	101	6.1	1819	2	A71928	cag island protein
15	101	6.1	3011	1	GNWVC3	genome polyprotein
16	100.5	6.1	633	2	T24898	hypothetical prote
17	100	6.1	395	2	S44454	transcription fact
18	100	6.1	520	2	G84510	probable receptor-
19	99.5	6.0	677	2	C42125	trophozoite cystei
20	99	6.0	373	2	T47488	hypothetical prote
21	99	6.0	565	2	T16408	hypothetical prote
22	99	6.0	2195	2	T34264	hypothetical prote
23	98.5	6.0	574	2	B88465	protein B0244.8 [i
24	98.5	6.0	962	2	JC5571	subtilisin-like pr
25	98.5	6.0	3106	1	S53868	laminin alpha-2 ch
26	98.5	6.0	5376	2	T42215	zonadhesin - mouse
27	97.5	5.9	432	2	T37509	hypothetical prote
28	97.5	5.9	969	1	A39490	subtilisin-like pr
29	97.5	5.9	975	2	JC5570	subtilisin-like pr

RESULT 1
T23152
hypothetical protein JC8.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T23152
R:Lightning, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19699
A:Accession: T23152
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-429 <WIL>
A:Cross-references: EMBL:Z82274; PIDN:CA805228.1; GSPDB:GN00022; CESP:JC8.6b
A:Experimental source: clone JC8
C:Genetics:
A:Gene: CESP:JC8.6b
A:Map position: 4
A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

ALIGNMENTS

Query Match 26.5%; Score 437; DB 2; Length 429;
Best Local Similarity 41.3%; Pred. No. 6.8e-28;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;
Qy 98 YCDCFASGDFCNCCNCCNHLHDIERFKAICACLRNPEAFQPKIG--KGQLGNVKP 155
Db 182 YCDCFANGFECDCKCKDCHNNIEYDSQRSRAIROSLERNPNAPFKIGIARGITDIER 241
Qy 156 QHNKGCNCRSGCLKNYCEYEAQIMCSSICKICCKN-----YEESE-----PERKTL 203
Db 242 LHQKGCHCKSGCLKNYCEYEAQVPCDRCCKCKCQNTETRYMTRYKNSGGAVSNTNAL 301
Qy 204 MSMPNMYMT-----GGL-----EGSHYLPPTKFGSLPRFSDHRR--PSSCISWEVVEAT 250
Db 302 MSLNASTATPDSGPGSVVTDEHGDY-EDMLLSHKPKVEMDPFRFPWYMTDEVVEAA 360
Qy 251 CACLLAQGEAA-----BKEHCCKCLAEQIMLEFGRCLSQL 287
Db 361 TCMVVAQAEALNVEKVTQDEKLNMEKLVLRFGRCLEQMI 403

RESULT 2

B84585
hypothetical protein At2g20110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84585
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999

trophozoite cystei
hypothetical prote
integrin beta-7 ch
genome polyprotein
Acorfi-70 protein -
hypothetical prote
trithorax protein
integrin beta-1 ch
genome polyprotein
laminin beta-2 cha
cag pathogenicity
genome polyprotein
integrin beta 2 ch
hypothetical prote
laminin alpha-2 ch
tenascin precursor

Db 110 ESPKARGPRPVEGRDGT PQKKKQCNCCKHSRCLKLYCECFASGTYCDGCNCVCNCFNNVDN 169

A; Reference number: Z20425

QY 265 HC-----SKCLA 271
 II
 Db 974 HCVSSSLQOCLS 986

RESULT 7
 IUHULM
 leukocyte adhesion protein beta chain (CD18) precursor - human
 N;Alternate names: integrin beta-2; LFA-1 beta chain; Mac-1 beta chain; p150,95 beta
 C;Species: Homo sapiens (man)
 C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 23-Jul-1999
 C;Accession: A25967; A91084; S19324; I70090; I70091; A29265; A34497
 R;Kishimoto, T.K.; O'Connor, K.; Lee, A.; Roberts, T.M.; Springer, T.A.
 Cell 48, 681-690, 1987
 A;Title: Cloning of the beta subunit of the leukocyte adhesion proteins: homology to
 A;Reference number: A25967; MUID:87131080
 A;Accession: A25967
 A;Molecule type: mRNA
 A;Residues: 1-769 <RIS>
 A;Cross-references: GB:M15395; NID:gl86933; PIDN:AAA59490.1; PID:g307113
 A;Note: source of LFA-1 was the SKW3 T-cell line; source of Mac-1 was pooled leukocyte
 R;Law, S.K.A.; Gagnon, J.; Hildreth, J.E.; Wells, C.E.; Willis, A.C.; Wong, A.J.
 EMBO J. 6, 915-919, 1987
 A;Title: The primary structure of the B-subunit of the cell surface adhesion glycopro
 A;Reference number: A91084; MUID:87246525
 A;Accession: A91084
 A;Molecule type: mRNA
 A;Residues: 9-198,'p',200-769 <LA2>
 A;Cross-references: GB:Y0057; NID:g30228; PIDN:CAA68266.1; PID:g762939
 R;Weitzman, J.B.; Wells, C.E.; Wright, A.H.; Clark, P.A.; Law, S.K.A.
 FEBS Lett. 294, 97-103, 1991
 A;Title: The gene organisation of the human beta2 integrin subunit (CD18).
 A;Reference number: S19324; MUID:92077153
 A;Accession: S19324
 A;Molecule type: DNA
 A;Residues: 1-3;16-23;46-53;106-113;163-170;244-251;296-303;328-335;358-365;405-412;4
 R;Nelson, C.; Rabb, H.; Arnaout, M.A.
 J. Biol. Chem. 267, 3351-3357, 1992
 A;Title: Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing a
 A;Reference number: I55376; MUID:92147694
 A;Accession: I70090
 A;Status: translated from GB/EMBL/DBDJ
 A;Molecule type: mRNA
 A;Residues: 347-350,'S',352-355 <NEL>
 A;Cross-references: GB:S81234; NID:g245294; PIDN:AAB21404.1; PID:g245295
 A;Note: mutant form
 A;Accession: I70091
 A;Status: preliminary; translated from GB/EMBL/DBDJ
 A;Molecule type: DNA
 A;Residues: 246-247,'PSSQ',248-249 <RES>
 A;Cross-references: GB:S81252; NID:g245289; PIDN:AAB21402.1; PID:g245290
 A;Note: mutant splice form
 C;Comment: The leukocyte adhesion proteins are noncovalently linked heterodimers of d
 and a deficiency of them is attributed to a genetic defect in the expression or stru
 C;Comment: The cysteine residues are involved in intrachain disulfide bonds.
 C;Genetics:
 A;Gene: GDB:ITGB2
 A;Cross-references: GDB:120574; OMIM:600065
 A;Map position: 21q22.3-21q22.3
 C;Superfamily: Integrin beta chain; laminin-type EGF-like homology
 C;Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; leuk
 F;1-22/Domain: signal sequence #status predicted <SIG>
 F;23-769/Product: leukocyte adhesion protein beta chain #status predicted <MAT>
 F;23-670/Domain: extracellular #status predicted <EXT>
 F;445-631/Region: cysteine-rich
 F;459-540/Region: duplication
 F;541-627/Region: duplication
 F;701-723/Domain: transmembrane #status predicted <MEM>
 F;724-769/Domain: intracellular #status predicted <CYT>
 F;23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic
 F;50,116,212,254,501/Binding site: carbohydrate (Asn) (covalent) #status predicted

Db 937 PVTQLQACVCDPGYIGSRCDDCASGFFGNPDSFGSCQPC--QCHNIDT-----T 985
QY 137 NPEAFQPKTKG-----GOLGNVYKQHNKGCNCRSGCLKNY-----C 173
Db 986 DPEACDKDGRCLKCLYHTTEGDHCLQCYGYGDDALRQDCRKCVCNLYLGTVKEHCNGSDC 1045
QY 174 EYEAQIMCSSTCKIGCKNYEESPERKTLMSPNYMQ-----TG-----GLESHYLPPT- 224
Db 1046 HCDKATGQCSCLPNVIG-----QNCDR-----CAPNTWLASGTGCGPCNCNAHSGFSPSC 1096
QY 225 -KFSG---LPRFSDHRRPSSC--ISW-----EVEATCACLLQ 257
Db 1097 NEFTGQCCMPGFG-GRTSCSEOLFWGDDVECRACDCDPRGIETPQDQSTGQCVVE 1155
QY 258 GEEAEKEHCKSC 269
Db 1156 GVEGPR--CDKC 1165
RESULT 12
MMFFBI
N:Alternate names: laminin chain B1
C:Species: Drosophila melanogaster
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C:Accession: A28783; S14462; B28783
R:Montell, D.J.; Goodman, C.S.
Cell 53, 463-473, 1988
A:Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals do
A:Reference number: A28783; MUID:88210471
A:Accession: A28783
A:Molecule type: mRNA
A:Residues: 1-1790 <MON1>
A:Cross-references: EMBL:M19525
R:Montell, D.J.; Goodman, C.S.
submitted to the EMBL Data Library, June 1988
A:Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reve
A:Reference number: S14462
A:Accession: S14462
A:Molecule type: mRNA
A:Residues: 1-667, 'L', 669-725, 'V', 728-947, 950-1790 <MON2>
A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
C:Genetics:
A:Gene: lamB1
A:Cross-references: FlyBase:FBgn0002527
A:Map position: 2L 28D
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C:Function:
A:Description: Interact with cells and with other basement membrane proteins to promote
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
F:27-288/Domain: VI <DOM6>
F:289-561/Domain: V <DOM5>
F:290-354/Domain: laminin-type EGF-like homology <LE01>
F:357-417/Domain: laminin-type EGF-like homology <LE02>
F:420-477/Domain: laminin-type EGF-like homology <LE03>
F:480-528/Domain: laminin-type EGF-like homology <LE04>
F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:562-789/Domain: IV <DOM4>
F:643-645/Region: cell attachment (R-G-D) motif
F:790-1189/Domain: III <DOM3>
F:791-836/Domain: laminin-type EGF-like homology <LE06>
F:839-882/Domain: laminin-type EGF-like homology <LE07>
F:885-932/Domain: laminin-type EGF-like homology <LE08>
F:935-990/Domain: laminin-type EGF-like homology <LE09>
F:968-972/Region: cell adhesion #status predicted
F:993-1042/Domain: laminin-type EGF-like homology <LE10>
F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
F:1190-1407/Domain: II <DOM2>

F:1408-1434/Domain: alpha <ALP>
F:1435-1790/Domain: I <DOM1>
F:51-56/Disulfide bonds: #status predicted
F:140,203,234,489,593,1053,1248,1303,1332,1343,1475,1517,1583,1646,1705/Binding
F:1191,1194,1788/Disulfide bonds: interchain #status predicted

Query Match 6.2%; Score 102; DB 1; Length 1790;

Best Local Similarity 21.4%; Pred. No. 3.8;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

QY 97 GYCD-----CF--ASGDFCNC-----NCCNC-----CNLHHDIERFKA 129
Db 1005 GNCDDRTGACLKCLYQTGDHCELCCKDGFGLALQNCQCECDFLGTNTTAHCDRTFG 1064

QY 130 IKACLGRLNPEAFQPKIGKQGLGNVYKQHNK-----GCNRRSGCLKNYCECEVAQIM 181
Db 1065 QCPCL-----PNVGVRCDOCAENHWKIASGEGCESCNDPIGALHEOCNSVTGQCQ 1116

QY 182 CS---SICKICGCK-NYEESEPERKTLMSPNYMQTGLGESHYLPPTKFS--GLPRFSDH 235
Db 1117 KPGFGGRACNOCAHYWGNPNK-----COPCECDQGAADFQCD 1157

QY 236 RRPSSCISWEVVEATCACLQAQEEAEKEHCKSLAEQMILEEFGRCLS 284

Db 1158 RETGNCVHEGIGGYKNECARGYICQFPHCSPC-----GECFN 1196

RESULT 13
S08167
Balbiani ring 3 protein - midge (Chironomus tentans)
C:Species: Chironomus tentans
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 21-Jul-2000
C:Accession: S08167
R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.
J. Mol. Biol. 211, 331-349, 1990
A:Title: The balbiani ring 3 gene in Chironomus tentans has a diverged repetitive str
A:Reference number: S08167; MUID:90172404
A:Accession: S08167
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1700 <PAU>
A:Cross-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058
C:Genetics:
A:Gene: BR3
A:Map position: 4
C:Superfamily: unassigned Balbiani ring proteins

Query Match 6.1%; Score 101; DB 2; Length 1700;

Best Local Similarity 22.6%; Pred. No. 4.3;
Matches 51; Conservative 20; Mismatches 87; Indels 68; Gaps 15;

QY 99 CDCFASGD-----FC--NCCNC-----NCCNLLHHDIERFKAICACILGRNPEAF 141
Db 198 CSCEKGDGKCGSKTWCKNCRICPTAEPAGCCSAPLKWDDD--KSCACAPAKWEEK 255

QY 142 QPKIGKQGLGNVYKQHNKGCNRRSGCLKN-----YCEYE-----AQIMCSSICK 187
Db 256 EKCVESGKIWN---PNTCEGCAQLNCPDNKANKETCQCEKEVKKCGGVFCKDSCS 312

QY 188 CIGCKNYEESPERKTLMSPNYMQTGLGESHYLP---PTKFSGLPR----- 231
Db 313 CV-CPGDKD---KTCTAPQVY---DGVACSCSPVNMQKPADGCGPRQKWKREKCEC 365

QY 232 -FSHRRPSSCISWEVVEATCAC-----LIAQGEAEKEHCS--KCL 270

Db 366 PVKHDCKNGKV--WDETICQICPRDPAVCTACKERGECSECKCI 409

RESULT 14
A71928
cag island protein - Helicobacter pylori (strain J99)

Result No.	Score	Query %			ID	Description
		Match	Length	DB		
1	1647	100.0	299	1	MT15_HUMAN	Q9Y415 homo sapien
2	1251	76.0	295	1	MT15_MOUSE	O9Wtj6 mus musculu
3	109.5	6.6	769	1	ITB2_HUMAN	P05107 homo sapien
4	106.5	6.3	3718	1	LMAS_MOUSE	Q61001 mus musculu
5	103.5	6.3	1786	1	LMB1_MOUSE	P02469 mus musculu
6	103	6.3	769	1	ITB2_PIG	P53714 sus scrofa
7	102	6.2	1790	1	LMB1_DROME	PL1046 drosophila
8	101.5	6.2	396	1	TFH2_MOUSE	Q9Y1b4 mus musculu
9	101	6.1	1700	1	BAR3_CHITE	Q03376 chiromomus
10	101	6.1	3011	1	POLG_HCV1	P26664 h genome po
11	100.5	6.1	760	1	EZ_DROME	P42124 drosophila
12	100	6.1	3695	1	LMAS5_HUMAN	Q15230 homo sapien
13	99	6.0	787	1	ITB6_MOUSE	Q9Z0t9 mus musculu
14	98.5	6.0	3106	1	LMB2_MOUSE	Q06075 mus musculu
15	98.5	6.0	5376	1	ZAN_MOUSE	O88799 mus musculu
16	97.5	5.9	328	1	C170_GIALA	P15799 giardia lam
17	97.5	5.9	969	1	PAC4_HUMAN	P29122 homo sapien
18	97	5.9	395	1	TFH2_HUMAN	Q13888 homo sapien
19	97	5.9	798	1	ITB7_HUMAN	P26010 homo sapien
20	97	5.9	3010	1	POLG_HCVJT	Q00269 h genome po
21	96.5	5.9	290	1	Y070_NPVAC	P41470 autographa
22	96.5	5.9	1895	1	YLK3_CAEEL	P41951 caenorhabdi
23	96.5	5.9	3726	1	TRX_DROME	P26659 drosophila
24	96	5.8	798	1	ITB1_XENLA	P12606 xenopus lae
25	96	5.8	1808	1	TENA_CHICK	P10039 gallus galli
26	96	5.8	3010	1	POLG_HCVBK	P26663 h genome po
27	95.5	5.8	747	1	E2H1_HUMAN	Q92800 homo sapien
28	95.5	5.8	1801	1	LMB2_RAT	P15800 rattus norv
29	94	5.7	3011	1	POLG_HCVH	P27958 h genome po
30	93	5.6	3110	1	LMAS2_HUMAN	P24043 homo sapien
31	93	5.6	3224	1	RBP2_HUMAN	P49792 homo sapien
32	92.5	5.6	332	1	XYNA_CRYAL	P07529 cryptococcu
33	92.5	5.6	611	1	LEM2_CANFA	P33730 canis famil

```
QY 181 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 240
Db 181 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 240

QY 241 CISWVVEATCACLLAQGEAEKHCSCLAEQMILEEFGRCLSQILHTEFKSKGLKME 299
Db 241 CISWVVEATCACLLAQGEAEKHCSCLAEQMILEEFGRCLSQILHTEFKSKGLKME 299

RESULT 2
MTL5_MOUSE STANDARD; PRT; 295 AA.
ID ITB2_MOUSE
AC P05107; 016418;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins IFA-
DE 1/CR3/p150.95 beta-subunit) (CD18) (Complement receptor C3 beta-
DE subunit).
GN ITGB2 OR CD18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RT early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
CC PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U77383; AAD24667.1; -
CC EMBL; U67176; AAD24666.1; -
CC MGD; MGI:1340029; Mtl5.
KW Spermatogenesis...
FT DOMAIN 99 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Query Match 76.0%; Score 1251; DB 1; Length 295;
Best Local Similarity 76.3%; Pred. No. 8.6e-101;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

QY 1 MVICQLAGGQMLCIDNSRTELKALHLVPOYODNNYLOSVDPKPMTALVGRFLPASTK 60
Db 1 MVICQLAGGQMLCIDNCGARELKHLLPQYDDQSFQPSSELPKPMITLVGLRLLPVPK 60

QY 61 LNLITQDNGALPSVNGSAFFSGSTLPGPKITLAGYCDCAFSGDFCNCCNCCNNL 120
Db 61 LNLITQDNGALPSAVNGAAPPSPALQGPPIKTLGYCDCFSSGDFCNCS- - -CNNL 116

QY 121 HDIERFKAIKACLRNPFQPKIGKQGLGNVQPKQINKCNCRRSCLKNKYCEYQAI 180
Db 117 RHELERFKAIKACLRNPFQPKMGKGRGLGAARLRSKGNCKRSCLKNKYCEYAKI 176

QY 181 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 240
Db 177 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 236
```

```
QY 241 CISWVVEATCACLLAQGEAEKHCSCLAEQMILEEFGRCLSQILHTEFKSKGLKME 299
Db 237 CISWVVEATCACLLAQGEAEKHCSPSLAEQMLEEFGRCLSQILHTEFKSKGLKIE 295

RESULT 3
ITB2_HUMAN STANDARD; PRT; 769 AA.
ID ITB2_HUMAN
AC P05107; 016418;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins IFA-
DE 1/CR3/p150.95 beta-subunit) (CD18) (Complement receptor C3 beta-
DE subunit).
GN ITGB2 OR CD18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87131080; PubMed=3028646;
RA Kishimoto T.K., O'Connor K., Lee A., Roberts T.M., Springer T.A.;
RT "Cloning of the beta subunit of the leukocyte adhesion proteins:
RT homology to an extracellular matrix receptor defines a novel
RT supergene family.";
RL Cell 48:681-690(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92077153; PubMed=1683838;
RA Weitzman J.B., Wells C.E., Wright A.H., Clark P.A., Law S.K.A.;
RT "The gene organisation of the human beta 2 integrin subunit (CD18).";
RL FEBS Lett. 294:97-103(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hatton M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Antonarakis S.E.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [4]
RP SEQUENCE OF 9-769 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Spleen;
RX MEDLINE=872476525; PubMed=2954816;
RA Law S.K.A., Gagnon J., Hildreth J.E., Wells C.E., Willis A.C.,
RA Wong A.J.;
RT "The primary structure of the beta-subunit of the cell surface
RT adhesion glycoproteins LFA-1, CR3 and p150,95 and its relationship to
RT the fibronectin receptor.";
RL EMBO J. 6:915-919(1987).
RN [5]
RP SEQUENCE OF 347-355 FROM N.A., AND VARIANTS LAD1 SER-351 AND TRP-586.
RX MEDLINE=92147694; PubMed=1346613;
RA Nelson C., Rabb H., Arnaout M.A.;
RT "Genetic cause of leukocyte adhesion molecule deficiency. Abnormal
RT splicing and a missense mutation in a conserved region of CD18 impair
RT cell surface expression of beta 2 integrins.";
RL J. Biol. Chem. 267:3351-3357(1992).
RN [6]
RP VARIANTS LAD1 PRO-149 AND ARG-169.
RX MEDLINE=90293696; PubMed=1694220;
RA Wardlaw A.J., Hibbs M.L., Stacker S.A., Springer T.A.;
```


QY 104 -SGDFCN-----CNCNNCCNLLHDIERF-----KAIKACLGRLNPEAFQPKIGKQGLGN 152
 Db 544 YNQGVGGPGRGLFCGKC -RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR---- 595
 QY 153 VRPQHNKGNCRRSGGLKNVCYCEYAIW-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLCQECPCSPCGKYSACAECLKFEKGP 638

RESULT 4
 LMA5_MOUSE STANDARD; PRT: 3718 AA.
 AC Q61001; Q9JHQ6;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-5 chain precursor.
 GN LMA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE OF 1-92 FROM N.A.
 RA Timpl R., Sasaki T.;
 RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
 chain".
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 84-3718 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=96081906; PubMed=7499364;
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
 expression in adult mouse tissues.";
 RL J. Biol. Chem. 270:28523-28526(1995).
 RN [3]

RA Miner J.H., Lewis R.M., Sanes J.R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
 ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
 KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
 AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ293593; CAB999255.1; -

DR EMBL; U37501; AAC53430.1; -;
 DR HSSP; P02468; ITLE.
 DR MGD; MGI:105382; Lama5.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00052; laminin_B; 1.
 DR Pfam; PF00053; laminin_EGF; 19.
 DR Pfam; PF00054; laminin_G; 2.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR ProDom; PD002082; LamNT; 1.
 DR ProDom; PD003031; Laminin_B; 1.
 DR SMART; SM00180; EGF_Lam; 17.
 DR SMART; SM00001; EGF_Like; 2.
 DR SMART; SM00281; Lamb; 1.
 DR SMART; SM00282; LamG; 5.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
 DR PROSITE; PS00025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 40
 FT CHAIN 41 3718
 FT DOMAIN 41 304
 FT DOMAIN 305 363
 FT DOMAIN 364 433
 FT DOMAIN 434 479
 FT DOMAIN 500 546
 FT DOMAIN 547 592
 FT DOMAIN 593 637
 FT DOMAIN 638 682
 FT DOMAIN 683 728
 FT DOMAIN 729 781
 FT DOMAIN 782 833
 FT DOMAIN 834 855
 FT DOMAIN 856 1442
 FT DOMAIN 1443 1488
 FT DOMAIN 1489 1532
 FT DOMAIN 1533 1581
 FT DOMAIN 1582 1632
 FT DOMAIN 1633 1642
 FT DOMAIN 1643 1831
 FT DOMAIN 1832 1864
 FT DOMAIN 1865 1914
 FT DOMAIN 1915 1970
 FT DOMAIN 1971 2024
 FT DOMAIN 2025 2071
 FT DOMAIN 2072 2118
 FT DOMAIN 2119 2168
 FT DOMAIN 2169 2735
 FT DOMAIN 2736 2933
 FT DOMAIN 2947 3119
 FT DOMAIN 3128 3296
 FT DOMAIN 3337 3511
 FT DOMAIN 3518 3689
 FT DOMAIN 2205 2257
 FT DOMAIN 2330 2464
 FT DOMAIN 2604 2621
 FT DOMAIN 2639 2705
 FT SITE 1723 1725
 FT SITE 1839 1841
 FT SITE 305 314
 FT DISULFID 307 327
 FT DISULFID 329 338
 FT DISULFID 341 361
 FT DISULFID 364 373
 FT DISULFID 366 398
 FT DISULFID 401 410
 LAMININ ALPHA-5 CHAIN.
 LAMININ N-TERMINAL (DOMAIN VI).
 LAMININ EGF-LIKE 1.
 LAMININ EGF-LIKE 2.
 LAMININ EGF-LIKE 3.
 LAMININ EGF-LIKE 4.
 LAMININ EGF-LIKE 5.
 LAMININ EGF-LIKE 6.
 LAMININ EGF-LIKE 7.
 LAMININ EGF-LIKE 8.
 LAMININ EGF-LIKE 9.
 LAMININ EGF-LIKE 10.
 LAMININ EGF-LIKE 11 (INCOMPLETE).
 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 LAMININ EGF-LIKE 12.
 LAMININ EGF-LIKE 13.
 LAMININ EGF-LIKE 14.
 LAMININ EGF-LIKE 15.
 LAMININ EGF-LIKE 16 (N-TERMINAL).
 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 LAMININ EGF-LIKE 16 (C-TERMINAL).
 LAMININ EGF-LIKE 17.
 LAMININ EGF-LIKE 18.
 LAMININ EGF-LIKE 19.
 LAMININ EGF-LIKE 20.
 LAMININ EGF-LIKE 21.
 LAMININ EGF-LIKE 22.
 DOMAIN II AND I.
 LAMININ G-LIKE 1.
 LAMININ G-LIKE 2.
 LAMININ G-LIKE 3.
 LAMININ G-LIKE 4.
 LAMININ G-LIKE 5.
 COILED COIL (POTENTIAL).
 COILED COIL (POTENTIAL).
 COILED COIL (POTENTIAL).
 COILED COIL (POTENTIAL).
 CELL ATTACHMENT SITE (POTENTIAL).
 CELL ATTACHMENT SITE (POTENTIAL).
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M15525; AAA39407.1; ALT_INIT.

DR EMBL; X05212; CAA28839.1; -

DR PIR; A26413; MMSB1.

DR HSP; P02468; IKLO.

DR MGD; MGI:96743; Lambl-1.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001886; LamNT.

DR InterPro; IPR002049; Laminin_EGF.

DR Pfam; PF00053; laminin_EGF; 13.

DR Pfam; PF00055; laminin_Nterm; 1.

DR PRINTS; PR00011; EGFLAMININ.

DR ProDom; PD002082; LamNT; 1.

DR SMART; SM00180; EGF_Lam; 11.

DR SMART; SM00136; LamNT; 1.

DR PROSITE; PS00022; EGF_1; 9.

DR PROSITE; PS01186; EGF_2; 2.

DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;

KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

FT SIGNAL 1 21

FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.

FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).

FT DOMAIN 271 334 LAMININ EGF-LIKE 1.

FT DOMAIN 335 397 LAMININ EGF-LIKE 2.

FT DOMAIN 398 457 LAMININ EGF-LIKE 3.

FT DOMAIN 458 509 LAMININ EGF-LIKE 4.

FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).

FT DOMAIN 541 772 LAMININ DOMAIN IV.

FT DOMAIN 773 820 LAMININ EGF-LIKE 6.

FT DOMAIN 821 866 LAMININ EGF-LIKE 7.

FT DOMAIN 867 916 LAMININ EGF-LIKE 8.

FT DOMAIN 917 975 LAMININ EGF-LIKE 9.

FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.

FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.

FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.

FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.

FT DOMAIN 1179 1398 DOMAIN II.

FT DOMAIN 1398 1430 DOMAIN ALPHA.

FT DOMAIN 1431 1786 COILED COIL (POTENTIAL).

FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).

FT DOMAIN 1368 1388 COILED COIL (POTENTIAL).

FT DOMAIN 1448 1778 BY SIMILARITY.

FT DISULFID 271 280 BY SIMILARITY.

FT DISULFID 273 298 BY SIMILARITY.

FT DISULFID 300 309 BY SIMILARITY.

FT DISULFID 312 332 BY SIMILARITY.

FT DISULFID 335 344 BY SIMILARITY.

FT DISULFID 337 362 BY SIMILARITY.

FT DISULFID 365 374 BY SIMILARITY.

FT DISULFID 377 395 BY SIMILARITY.

FT DISULFID 398 411 BY SIMILARITY.

FT DISULFID 400 426 BY SIMILARITY.

FT DISULFID 428 437 BY SIMILARITY.

FT DISULFID 440 455 BY SIMILARITY.

FT DISULFID 458 472 BY SIMILARITY.

FT DISULFID 460 479 BY SIMILARITY.

FT DISULFID 481 490 BY SIMILARITY.

FT DISULFID 493 507 BY SIMILARITY.

FT DISULFID 507 785 BY SIMILARITY.

FT DISULFID 773 792 BY SIMILARITY.

FT DISULFID 775 792 BY SIMILARITY.

FT DISULFID 794 803 BY SIMILARITY.

FT DISULFID 806 818 BY SIMILARITY.

FT DISULFID 821 833 BY SIMILARITY.

FT DISULFID 823 840 BY SIMILARITY.

FT DISULFID 842 851 BY SIMILARITY.

FT DISULFID 854 864 BY SIMILARITY.

FT DISULFID 867 876 BY SIMILARITY.

FT DISULFID 889 883 BY SIMILARITY.

FT DISULFID 886 895 BY SIMILARITY.

FT DISULFID 898 914 BY SIMILARITY.

FT DISULFID 917 933 BY SIMILARITY.

FT DISULFID 919 944 BY SIMILARITY.

FT DISULFID 946 955 BY SIMILARITY.

FT DISULFID 958 973 BY SIMILARITY.

FT DISULFID 976 990 BY SIMILARITY.

FT DISULFID 978 997 BY SIMILARITY.

FT DISULFID 1000 1009 BY SIMILARITY.

FT DISULFID 1012 1025 BY SIMILARITY.

FT DISULFID 1084 1096 BY SIMILARITY.

FT DISULFID 1086 1103 BY SIMILARITY.

FT DISULFID 1105 1114 BY SIMILARITY.

FT DISULFID 1117 1129 BY SIMILARITY.

FT DISULFID 1132 1144 BY SIMILARITY.

FT DISULFID 1134 1151 BY SIMILARITY.

FT DISULFID 1153 1162 BY SIMILARITY.

FT DISULFID 1165 1176 BY SIMILARITY.

FT DISULFID 1179 1179 INTERCHAIN (PROBABLE).

FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).

FT DISULFID 1785 1785 INTERCHAIN (PROBABLE).

FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1041 1041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1195 1195 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1279 1279 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1343 1343 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1533 1533 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1542 1542 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1643 1643 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 1531 1534 SGNA -> NEMP (IN REF. 2).

FT CONFLICT 1749 1749 D -> N (IN REF. 2).

SQ SEQUENCE 1786 AA; 196904 MW; 846671B7BF41A474 CRC64;

Query Match 6.3%; Score 103.5; DB 1; Length 1786;

Best Local Similarity 21.4%; Pred.No. 0.5;

Matches 54; Conservative 29; Mismatches 74; Indels 95; Gaps 15;

QY 90 PPKITLAGYDCFAFGDFCNMC-----NCNCCNHLHDIERFAKACGLR 136

DB 937 PVTQLACVCDPGYIGSRCDGAGFGNPSDFGSGQPC-QCHHNIDT-----T 985

QY 137 NPEAFOPKIGK-----GOLGNVQPHNKGNCRRSGGLKNY-----C 173

DB 986 DPEACDKDTGRCLKCLYHTEGDHCLQCYGYGDALRQDKRCVCNYLGTVKHCNGSDC 1045

QY 174 EYEAQIMCSSICKCKGCKNVEESPERKTLMSPNYMQ-----TG-----GLESHYLPPT- 224

DB 1046 HCDKATGQCCLPNVIG-----QNCDR-----CAPNTWQLASGTGCGPCNCAHISFGPSC 1096

QY 225 -KFSG-----LPRFSDRRPSSC--ISW-----EVVEATCACLAAQ 257

DB 1097 NEFTGQCQCHPFGF-GRTSCQCLFWGDPVECRACDCPRGIETPQCQDSTGQCVCVE 1155

QY 258 GEAEKEHCKSKC 269

DB 1156 GVEGPR--CDKC 1165

RESULT 6

ITB2_PIG STANDARD; PRT; 769 AA.

ID ITB2_PIG AC P53714;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE	Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CR3/PI150,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit).
DE	ITGB2 OR CD18.
GN	Sus scrofa (Pig).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OC	NCBI_TaxID=9823;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	Lee J.K., Schook L.B., Rutherford M.S.;
RA	"Molecular cloning and characterization of the porcine CD18 leukocyte adhesion molecule.";
RT	Xenotransplantation 3:222-230(1996).
RL	-1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2, ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2 RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND VCAM1.
CC	-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2 ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D. SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC	-1- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U13941; AAB16688.1; -.
DR	HSSP; P00743; IAP0.
DR	InterPro; IPR000561; EGF-like.
DR	InterPro; IPR002369; Integrin_B.
DR	InterPro; IPR001169; Integrin_beta_C.
DR	InterPro; IPR003659; PSI.
DR	InterPro; IPR002035; VWFA.
DR	Pfam; PF00362; Integrin_B; 1.
DR	PRINTS; PR01186; INTEGRINB.
DR	ProDom; PD001811; Integrin_B; 1.
DR	SMART; SM00187; INB; 1.
DR	SMART; SM00423; PSI; 1.
DR	SMART; SM00327; VWA; 1.
DR	PROSITE; PS00243; INTEGRIN_BETA; 3.
DR	PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR	PROSITE; PS01186; EGF_2; UNKNOWN_3.
KW	Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat; Signal.
FT	SIGNAL 1 22 BY SIMILARITY.
FT	CHAIN 23 769 INTEGRIN BETA-2.
FT	DOMAIN 23 700 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 701 723 POTENTIAL.
FT	DOMAIN 724 769 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 124 363 VWFA-LIKE.
FT	DOMAIN 449 617 4 CYSTEINE-RICH TANDEM REPEATS.
FT	REPEAT 449 496 I.
FT	REPEAT 497 540 II.
FT	REPEAT 541 581 III.
FT	REPEAT 582 617 IV.
FT	DISULFID 25 447 BY SIMILARITY.
FT	DISULFID 33 43 BY SIMILARITY.
FT	DISULFID 36 73 BY SIMILARITY.
FT	DISULFID 46 62 BY SIMILARITY.
FT	DISULFID 191 198 BY SIMILARITY.
FT	DISULFID 246 286 BY SIMILARITY.
FT	DISULFID 386 400 BY SIMILARITY.
FT	DISULFID 420 662 BY SIMILARITY.
FT	DISULFID 445 449 BY SIMILARITY.

RL DNA Cell Biol. 12:573-587(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88210471; PubMed=3365769;
RA Montell D.J., Goodman C.S.:
RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain
RT reveals domains of homology with mouse.";
RL Cell 53:463-473(1988).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahiker C., Davenport L.B., Davies P.,
RA de Pablo J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Munkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirski R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC ENBL; M95811; AAD19752.1; -.
CC ENBL; M19525; AAA28663.1; -.
CC ENBL; AB003618; AAF52563.1; -.
CC PIR; A28783; MMFFB1.
CC HSSP; P02468; 1TLE.
CC FlyBase; FBgn0002527; LanB1.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR002049; Laminin_EGF.
CC Pfam; PF00053; laminin_EGF; 13.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGF_LAMININ.
CC ProDom; PD002082; LamNT; 1.
CC SMART; SM00180; EGF_Lam; 11.
CC SMART; SM00001; EGF_Like; 1.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 10.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 12.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1790 LAMININ BETA-1 CHAIN.
FT DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 290 356 LAMININ EGF-LIKE 1.
FT DOMAIN 357 419 LAMININ EGF-LIKE 2.
FT DOMAIN 420 479 LAMININ EGF-LIKE 3.
FT DOMAIN 480 530 LAMININ EGF-LIKE 4.
FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
FT DOMAIN 562 789 LAMININ DOMAIN IV.
FT DOMAIN 791 838 LAMININ EGF-LIKE 6.
FT DOMAIN 839 884 LAMININ EGF-LIKE 7.
FT DOMAIN 885 934 LAMININ EGF-LIKE 8.
FT DOMAIN 935 992 LAMININ EGF-LIKE 9.
FT DOMAIN 993 1044 LAMININ EGF-LIKE 10.
FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.
FT DOMAIN 1096 1143 LAMININ EGF-LIKE 12.
FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
FT DOMAIN 1191 1407 DOMAIN II.
FT DOMAIN 1408 1434 DOMAIN ALPHA.
FT DOMAIN 1435 1790 DOMAIN I.
FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).
FT DOMAIN 1455 1507 COILED COIL (POTENTIAL).
FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).
FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).
FT DISULFID 290 299 BY SIMILARITY.
FT DISULFID 292 320 BY SIMILARITY.
FT DISULFID 322 331 BY SIMILARITY.
FT DISULFID 334 354 BY SIMILARITY.
FT DISULFID 357 366 BY SIMILARITY.
FT DISULFID 359 384 BY SIMILARITY.
FT DISULFID 387 396 BY SIMILARITY.
FT DISULFID 399 417 BY SIMILARITY.
FT DISULFID 420 433 BY SIMILARITY.
FT DISULFID 422 448 BY SIMILARITY.
FT DISULFID 450 459 BY SIMILARITY.
FT DISULFID 462 477 BY SIMILARITY.
FT DISULFID 480 493 BY SIMILARITY.
FT DISULFID 482 500 BY SIMILARITY.
FT DISULFID 502 511 BY SIMILARITY.
FT DISULFID 514 528 BY SIMILARITY.
FT DISULFID 591 803 BY SIMILARITY.
FT DISULFID 793 810 BY SIMILARITY.
FT DISULFID 812 821 BY SIMILARITY.
FT DISULFID 824 836 BY SIMILARITY.
FT DISULFID 839 851 BY SIMILARITY.
FT DISULFID 841 858 BY SIMILARITY.
FT DISULFID 860 869 BY SIMILARITY.
FT DISULFID 872 882 BY SIMILARITY.
FT DISULFID 885 894 BY SIMILARITY.

```
FT DISULFID 887 901 BY SIMILARITY.
FT DISULFID 904 913 BY SIMILARITY.
FT DISULFID 916 932 BY SIMILARITY.
FT DISULFID 935 951 BY SIMILARITY.
FT DISULFID 937 962 BY SIMILARITY.
FT DISULFID 964 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 993 1007 BY SIMILARITY.
FT DISULFID 1017 1026 BY SIMILARITY.
FT DISULFID 1029 1042 BY SIMILARITY.
FT DISULFID 1045 1059 BY SIMILARITY.
FT DISULFID 1047 1066 BY SIMILARITY.
FT DISULFID 1068 1077 BY SIMILARITY.
FT DISULFID 1080 1093 BY SIMILARITY.
FT DISULFID 1096 1108 BY SIMILARITY.
FT DISULFID 1098 1115 BY SIMILARITY.
FT DISULFID 1117 1126 BY SIMILARITY.
FT DISULFID 1129 1141 BY SIMILARITY.
FT DISULFID 1144 1156 BY SIMILARITY.
FT DISULFID 1146 1163 BY SIMILARITY.
FT DISULFID 1165 1174 BY SIMILARITY.
FT DISULFID 1177 1188 BY SIMILARITY.
FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1788 1788 INTERCHAIN (PROBABLE).
FT SITE 643 645 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.2%; Score 102; DB 1; Length 1790;
Best Local Similarity 21.4%; Pred. No. 0.67;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

Qy 97 GYCD-----CF--ASGDFCNC-----NCNC-----C>NNLHDIERFKA 129
Db 1005 GNCDRQTGACLKLCYQTGDCHELCCKDFFGDALQNCQCECDPLGNTNTIAHCDRETF 1064

Qy 130 IKACILGRNPEAFQPKIGKGLGNVYKPOHNK-----GCNCRSGCLKNKYCEYEAQIM 181
Db 1065 QCPCD-----PNVQGVRCDCQCAENHWKIASGEGCESCNDPIGALHEQCNSYTGOCQ 1116

Qy 182 CS---SICKICGCK-NYBESPERKTLMSMPNYMOTGGLESHYLPPTKFS--GLPRFSGHD 235
Db 1117 CKPGFGRACNOCAHYWGNPEK-----CQPCEDQFGAADFOCD 1157

Qy 236 RPPSSCISNVEVATCACLQAGEAEKHCSCLAEOIMLEEFGRCLUS 284
Db 1158 RETGNCVCHGIGYKCNECARGYIGQPPHGSPC-----GECFN 1196
```

```
RESULT 8
TFH2_MOUSE
ID TFH2_MOUSE STANDARD; PRT; 396 AA.
AC Q9JIB4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TFIIH basal transcription factor complex p44 subunit (Basic
DE transcription factor 2 44 kDa subunit) (BTF2-p44) (General
DE transcription factor IIH polypeptide 2).
GN GTF2H2 OR BTF2P44.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RN SEQUENCE FROM N.A.
RP MEDLINE=20414747; PubMed=10958627;
RX Endrizzi M.G., Hadinoto V., Gowney J.D., Miller W., Dietrich W.F.;
RA "Genomic sequence analysis of the mouse Naip gene array.";
RL Genome Res. 10:1095-1102(2000).
CC -!- FUNCTION: Component of the core-TFIIH basal transcription factor
CC involved in nucleotide excision repair (NER) of DNA and, when
CC complexed to CAK, in RNA transcription by RNA polymerase II. The
CC N-terminus interacts with and regulates XPD whereas an intact C-
CC terminus is required for a successful escape of RNAP II from the
CC promoter (By similarity).
CC -!- SUBUNIT: One of the six subunits forming the core-TFIIH basal
CC transcription factor. Interacts with XPB, XPD, p62 and p34 (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 C4-TYPE ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF242432; AAF82753.1; -.
CC MGD; MGI:1345669; Gtf2h2.
CC InterPro; IPR001841; Znf_ring.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00518; 2F_RING_1; FALSE_NEG.
CC PROSITE; PS50089; 2F_RING_2; FALSE_NEG.
CC Transcription regulation; DNA repair; Nuclear protein; Zinc-finger.
CC ZN_FING 292 309 C4-TYPE.
CC ZN_FING 346 383 RING-TYPE.
CC DOMAIN 242 245 POLY-SER.
CC SEQUENCE 396 AA; 44687 MW; 1BC57AFD464AA83F CRC64;

Query Match 6.2%; Score 101.5; DB 1; Length 396;
Best Local Similarity 21.8%; Pred. No. 0.15;
Matches 48; Conservative 27; Mismatches 72; Indels 73; Gaps 11;

Qy 2 VICOLKGTQMLCIDSRTRELKALHLPQYQDQNNYLQSDVPKPTALVGRFLPASTKL 61
Db 210 VLARTGTGYHVLDETHYKELLAAHHVSPPPASSSS-----ECSLRMGFPQHTA 260

Qy 62 NLITQOLEGALPSVYVNGSAFPSPGSLPQPKITLAGYCDCFASGDFCNCNCNNC----- 116
Db 261 SLSDQD---AKPS-FSMAHLDDNNSTPG--LTLGGY-----FCQCRAKYCELPVE 305

Qy 117 -----C>NNLHH-----DIERFKAKACILGRNPEAFQPKIGKGLGN 152
Db 306 CKICGLTLVSAPHLARSYHFLPDLDAFOEISLEEKYGERFCYGCQ-----GE 352

Qy 153 VKPOHNKGCN-CRRSGCLKNKYCEYEAQIMCSCICKICGC 191
Db 353 LRDQHVYVCTVCONVFCVD--CDVF----VHDSLHCPCGC 386

RESULT 9
BAR3_CHITE.
ID BAR3_CHITE STANDARD; PRT; 1700 AA.
AC Q03376;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Balbiani RING protein 3 precursor.
GN BR3.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
```



```

FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 3011;
Best Local Similarity 24.8%; Pred. No. 1.4;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAFP-SGSLPGPKITLAGYDCDFASGDCNCCNCCNNL-----HHDIERFRAI 130
DB 1424 SVIPTSQGVVVVATDALMTGY-----TGDFSDVIDCNTCVTQVDFSLDPTFTITLTP 1478
QY 131 KACLRNPEAFQPKIGKQLG---NVKQHNKGCNRRSGCLKN--YCEYEAQIMCSSI 185
DB 1479 ODAVSRTQR--RGRTRGKPGIYRVAPGE-----RPSGMFDDSSVLCGYDA----- 1523
QY 186 KCICCKNVEESPERKTLMSNPNYQOTGGL-----EG-----SHYLPPTKF 226
DB 1524 ---GCANYELTP-AETVRLRAYNMTFGLPVCQDHLFEWGVFTGLTHIDAHFLSQTKQ 1578
QY 227 SG--LPRFSDRRPSSCISWEVVEATCACLQAQGEAEKEHCKSL 270
DB 1579 SGENLP-----YLVAYQATVCARAQAAPPSPWDQWVKCL 1611

RESULT 11
EZ_DROME STANDARD; PRT; 760 AA.
ID E2_DROME
AC P42124;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enhancer of zeste protein.
GN E(2).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019309; PubMed=8413234;
RA Jones R.S., Gelbart W.M.;
RT "The Drosophila Polycomb-group gene Enhancer of zeste contains a
region with sequence similarity to trithorax.";
```

```

RL Mol. Cell. Biol. 13:6357-6366(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97164720; PubMed=9012527;
RA Carrington E.A., Jones R.S.;
RT "The Drosophila Enhancer of zeste gene encodes a chromosomal protein:
examination of wild-type and mutant protein distribution.";
RL Development 122:4073-4083(1996).
CC -!- FUNCTION: NEGATIVELY REGULATES THE SEGMENT IDENTITY GENES OF THE
ANTENNAPEDIA (ANT-C) AND BITHORAX (BX-C) GENE COMPLEXES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE EZ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00180; AAC46462.1; -.
DR Flybase; FBgn0000629; E(z).
DR InterPro; IPR001005; Myb_DNA_bind.
DR InterPro; IPR001214; SET.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00395; SANT; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS0280; SET; 1.
DR Transcription regulation; Nuclear protein; DNA-binding; Repressor;
KW Developmental protein.
FT DOMAIN 505 510 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 538 619 CYS-RICH.
FT DOMAIN 632 743 SET.
SQ SEQUENCE 760 AA; 86965 MW; 4A28A8EAD7968AC4 CRC64;

Query Match 6.1%; Score 100.5; DB 1; Length 760;
Best Local Similarity 20.3%; Pred. No. 0.36;
Matches 36; Conservative 23; Mismatches 63; Indels 55; Gaps 8;

QY 99 CDCFASGDFCNN-CNCCNCCNHLHDIERFKAICACLRNPEAFQPKIGKGLGNVKKQH 157
DB 549 CSCIQTFNCFKFCNCSDDQN-----RF----- 572
QY 158 NKGCCNRRSGCLKNVCEYEAQIMCS-SICKICGCKNTEESPERKTLMSNPNYQOTGGL 216
DB 573 -PGCRC-KAQCNTKQPCYLAVRECDPDLCCACGADQF-----KLTKITCKNVQVQGLH 625
QY 217 GSHYLPPTKFGSLPRFSDRRPSSCISWEVVEATCACLQAQ-----GEAEKEHCKS 267
DB 626 KHLMAPSDIAGWGLFKEGAQKN-----EFISEYCEIISQDEADRRGRKVDKYMCS 678

RESULT 12
LMA5_HUMAN STANDARD; PRT; 3695 AA.
ID LMA5_HUMAN
AC O15230; Q9H1P1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LMA5 OR KIA0533 OR KIAA1907.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
```

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clapp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivasalho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLeay K., McMurray A.A.,
RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RL "The DNA sequence and comparative analysis of human chromosome 20.";
RT Nature 414:865-871(2001).
RN [2]
RP SEQUENCE OF 197-1934 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
RL DNA Res. 8:179-187(2001).
RN [3]
RP SEQUENCE OF 2051-3695 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [4]
RP SEQUENCE OF 2743-3695 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97415425; PubMed=9271224;
RA Durkin M.E., Loebel F., Mattei M.-G., Gilpin B.J., Albrechtsen R.,
RA Wewer U.M.;
RT "Tissue-specific expression of the human laminin alpha5-chain, and
RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal
RT mouse chromosome 2 near the locus for the ragged (Ra) mutation.";
RL FEBS Lett. 411:296-300(1997).
RN [5]
RP EXPRESSION IN RETINA.
RX MEDLINE=20422761; PubMed=10964957;
RA Libby R.T., Champlaud M.-F., Claudepierre T., Xu Y., Gibbons E.P.,
RA Koch M., Burgeson R.E., Hunter D.D., Brunken W.J.;
RT "Laminin expression in adult and developing retinae: evidence of two
RT novel CNS laminins.";
RL J. Neurosci. 20:6517-6528(2000).
CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION AND ORGANIZATION
CC OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -!- SUBUNIT: LAMININ-15 COMPLEX IS AN HETEROPOLYMER COMPOSED OF THREE
CC DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG
CC AND THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, LUNG, KIDNEY, SKELETAL
CC MUSCLE, PANCREAS, RETINA AND PLACENTA. LITTLE OR NO EXPRESSION IN

CC BRAIN AND LIVER.
CC -!- DOMAIN: DOMAIN G IS GLOBULAR AND IS PART OF THE MAJOR CELL-BINDING
CC SITE LOCATED IN THE LONG ARM OF THE LAMININ HETEROPOLYMER.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL354836; CAC22309.1; ALT_SEQ.
CC EMBL: AB067494; BAB67800.1;
CC EMBL: AB011105; BAA25459.1;
CC EMBL: Z95636; CAB09137.1;
CC HSSP: P02468; IKLO.
CC MIM: 601033;
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002049; Laminin_EGF.
CC InterPro: IPR001791; Laminin_G.
CC Pfam: PF00053; laminin_EGF; 1.
CC Pfam: PF00054; laminin_G; 2.
CC SMART: SM00180; EGF_Lam; 2.
CC SMART: SM00282; LamG; 5.
CC PROSITE: PS00022; EGF_1; 19.
CC PROSITE: PS01186; EGF_2; 3.
CC PROSITE: PS01248; LAMININ_TYPE_EGF; 16.
CC PROSITE: PS50025; LAM_G_DOMAIN; 5.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL 1 35
CC FT CHAIN 36 3695 LAMININ ALPHA-5 CHAIN.
CC FT DOMAIN 36 299 LAMININ N-TERMINAL (DOMAIN VI).
CC FT DOMAIN 300 358 LAMININ EGF-LIKE 1.
CC FT DOMAIN 359 428 LAMININ EGF-LIKE 2.
CC FT DOMAIN 429 474 LAMININ EGF-LIKE 3.
CC FT DOMAIN 494 540 LAMININ EGF-LIKE 4.
CC FT DOMAIN 541 586 LAMININ EGF-LIKE 5.
CC FT DOMAIN 587 631 LAMININ EGF-LIKE 6.
CC FT DOMAIN 632 676 LAMININ EGF-LIKE 7.
CC FT DOMAIN 677 722 LAMININ EGF-LIKE 8.
CC FT DOMAIN 723 775 LAMININ EGF-LIKE 9.
CC FT DOMAIN 776 828 LAMININ EGF-LIKE 10.
CC FT DOMAIN 829 850 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC FT DOMAIN 851 1437 LAMININ DOMAIN IV 1 (DOMAIN IV B).
CC FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.
CC FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.
CC FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.
CC FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.
CC FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC FT DOMAIN 1638 1830 LAMININ DOMAIN IV 2 (DOMAIN IV A).
CC FT DOMAIN 1831 1863 LAMININ EGF-LIKE 17.
CC FT DOMAIN 1864 1912 LAMININ EGF-LIKE 18.
CC FT DOMAIN 1913 1968 LAMININ EGF-LIKE 19.
CC FT DOMAIN 1969 2022 LAMININ EGF-LIKE 20.
CC FT DOMAIN 2023 2069 LAMININ EGF-LIKE 21.
CC FT DOMAIN 2070 2116 LAMININ EGF-LIKE 22.
CC FT DOMAIN 2117 2166 LAMININ II AND I.
CC FT DOMAIN 2167 2735 LAMININ G-LIKE 1.
CC FT DOMAIN 2736 2929 LAMININ G-LIKE 2.
CC FT DOMAIN 2941 3115 LAMININ G-LIKE 3.
CC FT DOMAIN 3124 3292 LAMININ G-LIKE 4.
CC FT DOMAIN 3340 3513 LAMININ G-LIKE 5.
CC FT DOMAIN 3520 3692 COILED COIL (POTENTIAL).
CC FT DOMAIN 2203 2221 COILED COIL (POTENTIAL).
CC FT DOMAIN 2335 2466 COILED COIL (POTENTIAL).
CC FT DOMAIN 2510 2670 CELL ATTACHMENT SITE (POTENTIAL).
CC FT SITE 1722 1724

OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	(1)	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Kidney;	
RX	MEDLINE=20547423; PubMed=11095652;	
RA	Arend L.J., Smart A.M., Briggs J.P.;	
RT	"Mouse beta(6) integrin sequence, pattern of expression, and role in	
RT	kidney development."	
RL	J. Am. Soc. Nephrol. 11:2297-2305(2000).	
CC	- - FUNCTION: INTEGRIN ALPHA-V/BETA-6 IS A RECEPTOR FOR FIBRONECTIN	
CC	AND CYTOTACTIN. IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGANDS	
CC	(BY SIMILARITY).	
CC	- - SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-6	
CC	ASSOCIATES WITH ALPHA-V (BY SIMILARITY).	
CC	- - SUBCELLULAR LOCATION: Type I membrane protein.	
CC	- - SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.	
CC	- - SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
EMBL	AF115376; AAD17212.1; -	
MGI	96615; Itgb6.	
DR	InterPro; IPR000561; EGF-like.	
DR	InterPro; IPR002369; Integrin_B.	
DR	InterPro; IPR001169; Integrin_beta_C.	
DR	InterPro; IPR003659; PSI.	
DR	InterPro; IPR002035; VWFA.	
DR	Pfam; PF00362; Integrin_B; 1.	
DR	PRINTS; PR01186; INTEGRINB.	
DR	ProDom; PD001811; Integrin_B; 1.	
DR	SMART; SM00001; EGF_like; 1.	
DR	SMART; SM00187; INB; 1.	
DR	SMART; SM00423; PSI; 1.	
DR	SMART; SM00327; VWFA; 1.	
DR	PROSITE; PS00243; INTEGRIN_BETA; 2.	
DR	PROSITE; PS00022; EGF_1; UNKNOWN_2.	
DR	PROSITE; PS01186; EGF_2; UNKNOWN_1.	
KW	Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;	
KW	Repeat; Signal.	
FT	SIGNAL 1 21	POTENTIAL.
FT	CHAIN 22 787	INTEGRIN BETA-6.
FT	DOMAIN 22 706	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 707 729	POTENTIAL.
FT	DOMAIN 730 787	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 131 371	VWFA-LIKE.
FT	DOMAIN 456 619	4 CYSTEINE-RICH TANDEM REPEATS.
FT	REPEAT 456 501	I.
FT	REPEAT 502 543	II.
FT	REPEAT 544 582	III.
FT	REPEAT 583 619	IV.
FT	DISULFID 23 454	BY SIMILARITY.
FT	DISULFID 31 41	BY SIMILARITY.
FT	DISULFID 34 70	BY SIMILARITY.
FT	DISULFID 44 59	BY SIMILARITY.
FT	DISULFID 137 204	BY SIMILARITY.
FT	DISULFID 252 293	BY SIMILARITY.
FT	DISULFID 394 406	BY SIMILARITY.
FT	DISULFID 426 669	BY SIMILARITY.
FT	DISULFID 452 456	BY SIMILARITY.
FT	DISULFID 467 479	BY SIMILARITY.
FT	DISULFID 476 511	BY SIMILARITY.
FT	DISULFID 481 490	BY SIMILARITY.
FT	DISULFID 492 502	BY SIMILARITY.
FT	DISULFID 517 522	BY SIMILARITY.

CELL ATTACHMENT SITE (POTENTIAL).	1840
BY SIMILARITY.	309
BY SIMILARITY.	322
BY SIMILARITY.	333
BY SIMILARITY.	356
BY SIMILARITY.	368
BY SIMILARITY.	393
BY SIMILARITY.	405
BY SIMILARITY.	426
BY SIMILARITY.	440
BY SIMILARITY.	447
BY SIMILARITY.	458
BY SIMILARITY.	471
BY SIMILARITY.	506
BY SIMILARITY.	515
BY SIMILARITY.	526
BY SIMILARITY.	538
BY SIMILARITY.	553
BY SIMILARITY.	560
BY SIMILARITY.	571
BY SIMILARITY.	584
BY SIMILARITY.	599
BY SIMILARITY.	605
BY SIMILARITY.	616
BY SIMILARITY.	629
BY SIMILARITY.	644
BY SIMILARITY.	650
BY SIMILARITY.	661
BY SIMILARITY.	674
BY SIMILARITY.	689
BY SIMILARITY.	696
BY SIMILARITY.	707
BY SIMILARITY.	720
BY SIMILARITY.	1438
BY SIMILARITY.	1440
BY SIMILARITY.	1457
BY SIMILARITY.	1468
BY SIMILARITY.	1471
BY SIMILARITY.	1543

Query Match	6.18; Score 100; DB 1; Length 3695;
Best Local Similarity	21.88; Pred. No. 2.2;
Matches	57; Conservative 23; Mismatches 100; Indels 82; Gaps 14;

QY	93	ITLGYCDCAFAGGFC	-----NNC-----NNCCNLLHDIERFKA	KA-132
Db	294	ISIGRCVCHHADACDAKDP	DFRLQCTQHNTCGTCDRCCEFGNQ--QPHK	PATAN 351
QY	133	-----CLGR-----NPE-----AFQPKIGKQ	LGNY---KPOHKGKNCRRSGCL 169	
Db	352	SANECQSCNCGHATDCY	DPDVRRRASQSLDGTGQGGVCDICQHTTG	VNCER--CL 409
QY	170	KNYCECYEAQIMCSICK	ICGCKN-----YEESEPKTILMSPNY-----	MQTGGLEG 217
Db	410	PGFYRSPNHPDPSHVH	RRCCNCSDFDTGTCEDLTGR--CYCRPNF	SGERCDVCAEGFTG 467
QY	218	SHYLPPTKFSGLPRF	SHDRPSSCISWEVATCACLQAQGEAEK	KERCSKLAQEMILE 277
Db	468	FPSCYPT-----PSS	NDTRQVLPAGQIVNCDCSAAGTQGNACKD-----	P 510
QY	278	EFGRLS-----QILH	TEFKSKG 295	
Db	511	RVGRCLCKPNFGQTH	CELCAPG 532	

RESULT 13		
ITB6_MOUSE	STANDARD;	PRT: 787 AA.
ID	ITB6_MOUSE	
AC	Q920T9;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	01-MAR-2002 (Rel. 41, Last annotation update)	
DE	Integrin beta-6 precursor.	
CN	ITGB6.	

Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
TISSUE=Kidney;
MEDLINE=20547423; PubMed=11095652;
Arend L.J., Smart A.M., Briggs J.P.;
"Mouse beta(6) integrin sequence, pattern of expression, and role in
kidney development.";
J. Am. Soc. Nephrol. 11:2297-2305(2000).
-|- FUNCTION: INTEGRIN ALPHA-V/BETA-6 IS A RECEPTOR FOR FIBRONECTIN
AND CYTOTACTIN. IT RECOGNIZES THE SEQUENCE R-G-D IT ITS LIGANDS
(BY SIMILARITY).
-|- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-6
ASSOCIATES WITH ALPHA-V (BY SIMILARITY).
-|- SUBCELLULAR LOCATION: Type I membrane protein.
-|- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
-|- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF115376; AAD17212.1; -
MGD; MGI:96615; Itgb6.
InterPro; IPR000561; EGF-like.
InterPro; IPR002369; Integrin_B.
InterPro; IPR001169; Integrin_beta_C.
InterPro; IPR001659; PSI.
InterPro; IPR002035; VWFA.
Pfam; PF00362; Integrin_B; 1.
PRINTS; PR01186; INTEGRINB.
PRODOM; PD001811; Integrin_B; 1.
SMART; SM00001; EGF_like; 1.
SMART; SM00187; INB; 1.
SMART; SM00423; PSI; 1.
SMART; SM00327; VWFA; 1.
PROSITE; PS00243; INTEGRIN_BETA; 2.
PROSITE; PS00022; EGF_1; UNKNOWN_2.
PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
Repeat; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 787 INTEGRIN BETA-6.
FT DOMAIN 22 706 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 707 729 POTENTIAL.
FT DOMAIN 730 771 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 131 371 VWFA-LIKE.
FT DOMAIN 456 619 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 456 501 I.
FT REPEAT 502 543 II.
FT REPEAT 544 582 III.
FT REPEAT 583 619 IV.
FT DISULFID 23 454 BY SIMILARITY.
FT DISULFID 31 41 BY SIMILARITY.
FT DISULFID 34 70 BY SIMILARITY.
FT DISULFID 44 59 BY SIMILARITY.
FT DISULFID 197 204 BY SIMILARITY.
FT DISULFID 252 293 BY SIMILARITY.
FT DISULFID 394 406 BY SIMILARITY.
FT DISULFID 426 669 BY SIMILARITY.
FT DISULFID 452 456 BY SIMILARITY.
FT DISULFID 467 479 BY SIMILARITY.
FT DISULFID 476 511 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
FT DISULFID 492 502 BY SIMILARITY.
FT DISULFID 517 522 BY SIMILARITY.

J. Immunol. 151:1789-1801(1993).

[3]

SEQUENCE OF 64-281 FROM N.A.
MEDLINE=95179178; PubMed=7874173;
Xu H., Wu X.R., Wewer U.M., Engvall E.;
"Murine muscular dystrophy caused by a mutation in the laminin alpha
2 (Lama2) gene.";
Nat. Genet. 8:297-302(1994).

[4]

X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
MEDLINE=20085745; PubMed=10619025;
Hohenester E., Tisi D., Talts J.F., Timpl R.;
"The crystal structure of a laminin G-like module reveals the
molecular basis of alpha-dystroglycan binding to laminins, perlecan,
and agrin.";
Mol. Cell 4:783-792(1999).

-!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT. BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

-!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-
4 (S-MEROSIN).

-!- SUBCELLULAR LOCATION: Extracellular.

-!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
COMPONENT).

-!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

-!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBALAR.

-!- DISEASE: DEFECTS IN LAMA2 ARE A CAUSE OF MURINE MUSCULAR DYSTROPHY
(DY2J).

-!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

-!- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U12147; AAC52165.1; -;
EMBL; X69869; CAA49502.1; -;
EMBL; S75315; AAB33573.1; -;
PDB; 1Q0U; 03-DEC-99.
MGD; MGI:99912; Lama2,
InterPro; IPR000561; EGF-like.
InterPro; IPR001886; LamNT.
InterPro; IPR000034; Laminin_B.
InterPro; IPR002049; Laminin_EGF.
InterPro; IPR001791; Laminin_G.
Pfam; PF00052; laminin_B_2.
Pfam; PF00053; laminin_EGF; 15.
Pfam; PF00054; laminin_G; 5.
Pfam; PF00055; laminin_Nterm; 1.
PRINTS; PR00011; EGFLAMININ.
ProDom; PD002082; LamNT; 1.
ProDom; PD003031; Laminin_B; 2.
SMART; SM00180; EGF_Lam; 15.
SMART; SM00001; EGF_Like; 1.
SMART; SM00281; LamB; 2.
SMART; SM00282; LamG; 5.
SMART; SM00136; LamNT; 1.
PROSITE; PS00022; EGF_1; 11.
PROSITE; PS01186; EGF_2; 3.
PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
PROSITE; PS50025; LAM_G_DOMAIN; 5.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.

```
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 3106 LAMININ ALPHA-2 CHAIN.
FT DOMAIN 23 282 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 283 339 LAMININ EGF-LIKE 1.
FT DOMAIN 340 409 LAMININ EGF-LIKE 2.
FT DOMAIN 410 464 LAMININ EGF-LIKE 3.
FT DOMAIN 465 513 LAMININ EGF-LIKE 4.
FT DOMAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 720 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.
FT DOMAIN 1056 1101 LAMININ EGF-LIKE 12.
FT DOMAIN 1102 1161 LAMININ EGF-LIKE 13.
FT DOMAIN 1162 1171 LAMININ EGF-LIKE 14 (N-TERMINAL).
FT DOMAIN 1172 1375 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1376 1415 LAMININ EGF-LIKE 14 (C-TERMINAL).
FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15.
FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.
FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.
FT DOMAIN 1570 2140 DOMAIN II AND I.
FT DOMAIN 2141 2324 LAMININ G-LIKE 1.
FT DOMAIN 2325 2517 LAMININ G-LIKE 2.
FT DOMAIN 2522 2706 LAMININ G-LIKE 3.
FT DOMAIN 2707 2930 LAMININ G-LIKE 4.
FT DOMAIN 2929 3106 LAMININ G-LIKE 5.
FT DOMAIN 1662 1863 COILED COIL (POTENTIAL).
FT DOMAIN 1823 2146 COILED COIL (POTENTIAL).
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 285 303 BY SIMILARITY.
FT DISULFID 305 314 BY SIMILARITY.
FT DISULFID 317 337 BY SIMILARITY.
FT DISULFID 340 349 BY SIMILARITY.
FT DISULFID 342 374 BY SIMILARITY.
FT DISULFID 377 386 BY SIMILARITY.
FT DISULFID 389 407 BY SIMILARITY.
FT DISULFID 410 422 BY SIMILARITY.
FT DISULFID 412 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 452 462 BY SIMILARITY.
FT DISULFID 465 478 BY SIMILARITY.
FT DISULFID 467 482 BY SIMILARITY.
FT DISULFID 484 493 BY SIMILARITY.
FT DISULFID 496 511 BY SIMILARITY.
FT DISULFID 753 762 BY SIMILARITY.
FT DISULFID 755 769 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 784 800 BY SIMILARITY.
FT DISULFID 803 818 BY SIMILARITY.
FT DISULFID 805 828 BY SIMILARITY.
FT DISULFID 831 840 BY SIMILARITY.
FT DISULFID 843 858 BY SIMILARITY.
FT DISULFID 861 875 BY SIMILARITY.
FT DISULFID 882 882 BY SIMILARITY.
FT DISULFID 885 894 BY SIMILARITY.
FT DISULFID 897 911 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 916 933 BY SIMILARITY.
FT DISULFID 935 944 BY SIMILARITY.
FT DISULFID 947 960 BY SIMILARITY.
FT DISULFID 963 975 BY SIMILARITY.
FT DISULFID 965 981 BY SIMILARITY.
FT DISULFID 983 992 BY SIMILARITY.
FT DISULFID 995 1007 BY SIMILARITY.
FT DISULFID 1010 1019 BY SIMILARITY.
FT DISULFID 1012 1026 BY SIMILARITY.
FT DISULFID 1028 1037 BY SIMILARITY.

Query Match 6.0%; Score 98.5; DB 1; Length 3106;
Best Local Similarity 22.0%; Pred. No. 2.4;
Matches 70; Conservative 21; Mismatches 82; Indels 145; Gaps 19;

Qy 73 PSVNGSAFP-----SGSTL---PGPPKITLAGYDCDFAS---GDPC 108
Db 851 PVPGGSCQPCQNDNLDSIPGSCDSLSGCLCKPG---TTGRYCELCAFGYFGDAV 906
Qy 109 NNCNCNNC-CN-----NLHHDIERFKAIKACLRN-----PEAFQPKIKGQGLGNVKP 155
Db 907 NTKNCQPCRCIDINGSFSDCHTRTQCCECPNVQGRHCHDECKPETFGLQGRGCL----- 961
Qy 156 QHNKGCNRRSGCLKNYCEYEAIQMS----SICKICGCK-----NYEE----- 196
Db 962 ---PCNCSFGSKSFDCEA-SGQCWCQPGVAGKCKDCRCAHGYNFQBGCGCIACDCSHLG 1016
Qy 197 ---SPERKTLMSPNYMTGLEGSHYLPPT----- 224
Db 1017 NNCDPKTQCICPPN---TTGEKCECPNTWGHISIVTGCKVCNCSTVGSLASQCNCNVNTG 1073
Qy 225 ---KFSGLPRFSDRRPSSCI---SWEV-----VEAT-----CACLLA 256
Db 1074 QCSCHPFSGM-----KCECSRGHNYPLCTLCDCFLPCTDATTCDLTRKCSQSDQ 1126
Qy 257 QGE-----EAEKEHCKSC 269
Db 1127 TQCCKVNVGVHCDCR 1144

RESULT 15
ZAN_MOUSE
ID ZAN_MOUSE STANDARD; PRT; 5376 AA.
AC O88799; O08647;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98123114; PubMed=9452463;
```

RA Gao Z., Garbers D.L.;
 RT "Species diversity in the structure of zonadhesin, a sperm-specific
 RT membrane protein containing multiple cell adhesion molecule-like
 RT domains";
 RL J. Biol. Chem. 273:3415-3421(1998).
 RN [2]
 RP SEQUENCE OF 4864-5376 FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=97271566; PubMed=9126492;
 RA Gao Z., Harumi T., Garbers D.L.;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD.
 CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOEA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 25 VWFD DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U97068; AAC26680.1; -;
 DR EMBL; U83190; AAC53125.1; -;
 DR MGD; MGI:106656; ZAN.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR003645; FOLN.
 DR InterPro; IPR000998; MAM.
 DR InterPro; IPR002919; TIL.
 DR InterPro; IPR003328; TILA.
 DR InterPro; IPR001007; VWFC.
 DR InterPro; IPR001846; VWFD.
 DR Pfam; PF00629; MAM; 3.
 DR Pfam; PF01826; TIL; 25.
 DR Pfam; PF02345; TILA; 25.
 DR Pfam; PF00094; vwd; 4.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00274; FOLN; 11.
 DR SMART; SM00137; MAM; 2.
 DR SMART; SM00214; VWC; 17.
 DR SMART; SM00216; VWFD; 4.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 18.
 DR PROSITE; PS00740; MAM_1; FALSE_NEG.
 DR PROSITE; PS00060; MAM_2; 3.
 KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
 Repeat.
 FT SIGNAL
 FT CHAIN 1 17 POTENTIAL.
 FT DOMAIN 18 5376 ZONADHESIN.
 FT TRANSMEM 5311 5376 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 5338 5376 POTENTIAL.
 FT DOMAIN 45 210 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 215 374 MAM 1.
 FT DOMAIN 377 542 MAM 2.
 FT DOMAIN 547 1170 MAM 3.
 FT 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)

FT	DOMAIN	1171	1280	(MUCIN-LIKE DOMAIN).
FT	DOMAIN	1281	1669	VWFD 1 (PARTIAL).
FT	DOMAIN	1670	2056	VWFD 2.
FT	DOMAIN	2057	2459	VWFD 3.
FT	DOMAIN	2460	2579	VWFD 4.
FT	DOMAIN	2580	2699	VWFD 5 (PARTIAL).
FT	DOMAIN	2700	2819	VWFD 6 (PARTIAL).
FT	DOMAIN	2820	2939	VWFD 7 (PARTIAL).
FT	DOMAIN	2940	3059	VWFD 8 (PARTIAL).
FT	DOMAIN	3060	3179	VWFD 9 (PARTIAL).
FT	DOMAIN	3180	3299	VWFD 10 (PARTIAL).
FT	DOMAIN	3300	3416	VWFD 11 (PARTIAL).
FT	DOMAIN	3417	3536	VWFD 12 (PARTIAL).
FT	DOMAIN	3537	3656	VWFD 13 (PARTIAL).
FT	DOMAIN	3657	3776	VWFD 14 (PARTIAL).
FT	DOMAIN	3777	3892	VWFD 15 (PARTIAL).
FT	DOMAIN	3893	4928	VWFD 16 (PARTIAL).
FT	DOMAIN	4029	4148	VWFD 17 (PARTIAL).
FT	DOMAIN	4149	4263	VWFD 18 (PARTIAL).
FT	DOMAIN	4264	4283	VWFD 19 (PARTIAL).
FT	DOMAIN	4384	4503	VWFD 20 (PARTIAL).
FT	DOMAIN	4504	4623	VWFD 21 (PARTIAL).
FT	DOMAIN	4624	4743	VWFD 22 (PARTIAL).
FT	DOMAIN	4744	4863	VWFD 23 (PARTIAL).
FT	DOMAIN	4864	5261	VWFD 24 (PARTIAL).
FT	DOMAIN	5259	5295	VWFD 25.
FT	DISULFID	5263	5274	EGF-LIKE.
FT	DISULFID	5268	5283	BY SIMILARITY.
FT	DISULFID	5285	5294	BY SIMILARITY.
FT	CARBOHYD	339	339	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	499	499	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1216	1216	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1239	1239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1314	1314	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1814	1814	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1908	1908	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1933	1933	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2028	2028	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2111	2111	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2142	2142	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2332	2332	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2533	2533	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2575	2575	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2692	2692	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2812	2812	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3065	3065	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3144	3144	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3172	3172	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3288	3288	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3292	3292	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3782	3782	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4005	4005	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4136	4136	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4243	4243	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4254	4254	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4335	4335	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4376	4376	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4586	4586	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	5136	5136	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	5252	5252	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SEQUENCE	5376	5376	AA; 579908 MW; 0B44DB77DF2A2620 CRC64;
SQ				

Query Match 6.0%; Score 98.5; DB 1; Length 5376;
 Best Local Similarity 23.2%; Pred. No. 4.4;
 Matches 53; Conservative 19; Mismatches 73; Indels 83; Gaps 15;
 QY 99 CDCFASGDFCNCCNC-----NNCCNNLHDIERFKAIKACL-GRNPEAFQPKIGKQL 150
 DB 4232 CACVGGVQCHNFTCTPTGTQCNSSCSKITVQCPAHSQYTTCLPCLSPCDFP---EGLC 4288

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:22:18 ; Search time 92.59 Seconds
(without alignments)
558.651 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDSRT.....GRCLSQLHTEFKSKGLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	28.5	950	5 Q9V608	Q9v6q8 drosophila
2	437	26.5	429	5 O62295	O62295 caenorhabdi
3	437	26.5	435	5 Q95Q07	Q95qd7 caenorhabdi
4	376	22.8	571	10 Q9SL70	Q9sl70 arabidopsis
5	369	22.4	277	11 Q9D571	Q9d571 mus musculus
6	357	21.7	603	10 Q9SD1	Q9szd1 arabidopsis
7	262.5	15.9	356	10 Q9LW71	Q9lw71 arabidopsis
8	261	15.8	601	10 Q9CAV1	Q9cav1 arabidopsis
9	248.5	15.1	896	10 Q9ZS22	Q9zs22 glycine max
10	247.5	15.0	609	10 Q9M679	Q9m679 arabidopsis
11	245.5	14.9	658	10 Q23333	Q23333 arabidopsis
12	229.5	13.9	526	10 Q94A12	Q94a12 arabidopsis
13	229.5	13.9	695	10 Q9LE32	Q9le32 arabidopsis
14	229.5	13.9	695	10 Q9LUI3	Q9lui3 arabidopsis
15	202.5	12.3	593	10 Q9LUI5	Q9lui5 arabidopsis
16	200	12.1	243	5 Q9VMQ3	Q9vmq3 drosophila

17	187.5	11.4	553	10 Q94DS2	Q94ds2 oryza sativ
18	117	7.1	497	5 Q23460	Q23460 caenorhabdi
19	113	6.9	1981	5 Q9VVK7	Q9vvk7 drosophila
20	112	6.8	1737	4 Q75097	Q75097 homo sapien
21	107.5	6.5	1282	5 O18720	O18720 entamoeba d
22	106	6.4	3010	12 Q9DTE7	Q9dte7 hepatitis c
23	105.5	6.4	911	11 Q9CRX6	Q9crx6 mus musculu
24	105	6.4	984	5 Q9Y1P7	Q9y1p7 cryptospori
25	104	6.3	773	5 O17514	O17514 caenorhabdi
26	104	6.3	773	5 O62335	O62335 caenorhabdi
27	103.5	6.3	398	11 Q9LYN8	Q9lyn8 mus musculu
28	103.5	6.3	1125	5 P92135	P92135 entamoeba d
29	103.5	6.3	3010	12 Q9QIX2	Q9qix2 hepatitis c
30	103	6.3	552	15 Q83934	Q83934 ovine lenti
31	102.5	6.2	4123	4 Q75851	Q75851 homo sapien
32	102	6.2	420	5 P91776	P91776 pacifastacu
33	102	6.2	552	15 Q83935	Q83935 ovine lenti
34	102	6.2	2192	5 O01768	O01768 caenorhabdi
35	102	6.2	2327	13 Q9IBG7	Q9ibg7 xenopus lae
36	101	6.1	552	15 Q83932	Q83932 ovine lenti
37	101	6.1	552	15 Q83933	Q83933 ovine lenti
38	101	6.1	1819	16 Q9ZLV0	Q9zlv0 helicobacte
39	101	6.1	2436	12 Q81756	Q81756 hepatitis c
40	101	6.1	3010	12 Q9QIX8	Q9qix8 hepatitis c
41	101	6.1	3010	12 Q9QIX7	Q9qix7 hepatitis c
42	101	6.1	3011	12 Q9IFE5	Q9ife5 hepatitis c
43	100.5	6.1	633	5 Q22468	Q22468 caenorhabdi
44	100.5	6.1	760	5 Q9VTA3	Q9vta3 drosophila
45	100.5	6.1	1028	11 Q9JLL0	Q9jll0 mus musculu

ALIGNMENTS

RESULT 1

Q9V6Q8	PRELIMINARY;	PRT;	950 AA.
ID	Q9V6Q8		
AC	Q9V6Q8:		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	CG6061 PROTEIN.		
GN	CG6061.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Hough J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		

Best Local Similarity 29.7%; Pred. No. 6.7e-19;
Matches 69; Conservative 25; Mismatches 64; Indels 74; Gaps 10;

QY 33 ODONNYLOSD-----VPKPM TAL-----VGRFLPASTKLNLTQLEGAL-PSV----- 75
DB 307 RSTNDLPDSTSIKAPSPONCLDTSKQDTEILPRTIGL-----HLNGFVNPSVSSGR 363
QY 76 -----VNGSAPFSGS-----TLPGPKITLAG----- 97
DB 364 KKKIKDQAFPTTFHYNIEDEFTPVSTKRDLVFSDVKIMEPPERSVEGECDQLMA 423
QY 98 ----YDCFCASGDFCNN-CNCNCCNNLHHDIERFKAICACLRNPEAFQPKIGKQ--- 149
DB 424 MENRYCEFSAGLFCGEGPCSCQCNCFKPIHEDLVNKSREVIKARNPLAFAPKVVSTSDTV 483
QY 150 ----LGNVK-----POHNGKCNCRSGCKLNKCYCEYEAQIMCSSICCKIGCKN 193
DB 484 IDLWVNSKTPASARHKRGCKNRKSGCKKYCECFMGMVGCCSNCRCMGCKN 535

RESULT 9
ID Q92S22 PRELIMINARY; PRT; 896 AA.
AC Q92S22;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYSTEINE-RICH POLYCOMB-LIKE PROTEIN.
GN CPPL.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20345129; PubMed=10859345;
RA Cvitanič C., Pallisgaard N., Nielsen K.A., Hansen A.C., Larsen K.,
RA Pihakaski-Maunsbach K., Marcher K.A., Jensen E.O.;
RT "CPPL, a novel type DNA-binding protein involved in the expression of
RT a soybean leghemoglobin c3 gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:8163-8168(2000).
DR EMBL: AJ010165; CAA09028.1; -
SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;

Query Match 15.1%; Score 248.5; DB 10; Length 896;
Best Local Similarity 40.7%; Pred. No. 2.3e-17;
Matches 50; Conservative 13; Mismatches 31; Indels 29; Gaps 5;

QY 98 YDCFCASGDFCNN-CNCNCCNNLHHDIERFKAICACLRNPEAFQPKIGKQGNVKP 155
DB 487 YDCFAAGTCTDPCACGGCLNRPY-VETVETKQQTESRNPATAFAPKI-----VQP 538
QY 156 -----OHNGKCNRRSGCKLNKCYCEYEAQIMCSSICCKIGCKNVEE 196
DB 539 TTDISSHMDENLTTPSSARHKRGCKNRKSNCKLKYCECYQANVCCSGRCGCKNVHG 598
QY 197 SPE 199
DB 599 KKE 601

RESULT 10
ID Q9M679 PRELIMINARY; PRT; 609 AA.
AC Q9M679;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CXC DOMAIN CONTAINING TSO1-LIKE PROTEIN 1.
GN SOL1.
OS Arabidopsis thaliana (Mouse-ear cross).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LER;
RX MEDLINE=2033842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;
RT "TSOL is a novel protein that modulates cytokinesis and cell expansion
RT in Arabidopsis.";
RL Development 127:2219-2226(2000).
DR EMBL: AF205142; AAF69125.1; -
SQ SEQUENCE 609 AA; 66738 MW; 7E29C055A2423089 CRC64;

Query Match 15.0%; Score 247.5; DB 10; Length 609;
Best Local Similarity 38.6%; Pred. No. 1.8e-17;
Matches 49; Conservative 20; Mismatches 35; Indels 23; Gaps 6;

QY 98 YDCFCASGDFC-NNCNCCNN-LHHDIERFKAICACLRNPEAFQPKIGK----- 147
DB 341 YCEFAAGFYCIETPCSCNCFNKPPIHKDVV-LATRKQTESRNPAPAPKVRNSDSIIEV 399
QY 148 GOLGNVKP---QHNGKCNRRSGCKLNKCYCEYEAQIMCSSICCKIGCKN-----YE 195
DB 400 GEDASKTPASARHKRGCKNRKSNCKLKYCECYQGGVGSINCRCEGCKNAFGRKDSUFE 459
QY 196 ESPERKT 202
DB 460 QDEENET 466

RESULT 11
ID O23333 PRELIMINARY; PRT; 658 AA.
AC O23333;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 72.1 KDA PROTEIN.
GN AT4G14770.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;
RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Weitzenegger T., Pohl T.M., Terry N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kottler P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansoorge W., Cooke R., Berger C.,
RA Delsen M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chalvatzis N.
RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT Arabidopsis thaliana.";
RL Nature 391:485-488(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z97337; CAB10256.1; -
DR EMBL: AL161539; CAB78519.1; -
KW Hypothetical protein.

T-1-6 PETER WINTERGARTEN TUNING COURTESY THE TROOP OF THE 17TH CENT.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:16:00 ; Search time 34.45 seconds
(without alignments)
211.996 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSOILHTEFKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA:*
- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109.5	6.6	769	2 US-08-789-078-1	Sequence 1, Appl
2	109.5	6.6	769	2 US-08-752-633-1	Sequence 1, Appl
3	109.5	6.6	769	2 US-08-476-062A-45	Sequence 45, Appl
4	109.5	6.6	769	2 US-07-728-215-31	Sequence 31, Appl
5	109.5	6.6	769	4 US-08-938-085A-31	Sequence 31, Appl
6	109.5	6.6	769	5 PCT-US95-04886-1	Sequence 1, Appl
7	109.5	6.6	769	5 PCT-US96-01314-45	Sequence 45, Appl
8	108	6.6	676	3 US-08-630-172-10	Sequence 10, Appl
9	108	6.6	676	4 US-09-375-419-10	Sequence 10, Appl
10	108	6.6	739	4 US-08-444-818-148	Sequence 148, Appl
11	108	6.6	2995	4 US-08-444-818-138	Sequence 138, App
12	101	6.1	465	2 US-08-833-678A-2	Sequence 2, Appl
13	101	6.1	465	4 US-08-529-169A-2	Sequence 2, Appl
14	101	6.1	631	2 US-08-833-678A-1	Sequence 1, Appl
15	101	6.1	631	4 US-08-529-169A-1	Sequence 1, Appl
16	101	6.1	1021	1 US-07-910-760-12	Sequence 12, Appl
17	101	6.1	1021	1 US-08-440-519-12	Sequence 12, Appl
18	101	6.1	1021	4 US-08-440-549-12	Sequence 12, Appl
19	101	6.1	1786	4 US-08-444-818-54	Sequence 54, Appl
20	101	6.1	2261	4 US-08-444-818-66	Sequence 66, Appl
21	101	6.1	2436	4 US-08-444-818-75	Sequence 75, Appl
22	101	6.1	2772	4 US-08-444-818-75	Sequence 89, Appl
23	101	6.1	2894	2 US-08-466-975A-23	Sequence 23, Appl
24	101	6.1	2894	2 US-08-391-671A-23	Sequence 23, Appl
25	101	6.1	2894	3 US-08-467-902A-23	Sequence 23, Appl
26	101	6.1	2894	4 US-09-275-265-23	Sequence 23, Appl
27	101	6.1	3011	1 US-08-440-103-36	Sequence 36, Appl

28	101	6.1	3011	1 US-08-440-542-36	Sequence 36, Appl
29	101	6.1	3011	1 US-07-910-760-10	Sequence 10, Appl
30	101	6.1	3011	1 US-08-440-519-10	Sequence 10, Appl
31	101	6.1	3011	1 US-08-231-368-36	Sequence 36, Appl
32	101	6.1	3011	1 US-08-440-210-36	Sequence 36, Appl
33	101	6.1	3011	4 US-09-388-874-2	Sequence 2, Appl
34	101	6.1	3011	4 US-09-046-604-36	Sequence 36, Appl
35	101	6.1	3011	4 US-08-440-549-10	Sequence 10, Appl
36	97.5	5.9	969	2 US-08-284-941-2	Sequence 2, Appl
37	97.5	5.9	969	2 US-08-447-642-2	Sequence 2, Appl
38	97.5	5.9	969	4 US-09-236-503-2	Sequence 2, Appl
39	97.5	5.9	969	5 PCT-US93-02147A-2	Sequence 2, Appl
40	97	5.9	2955	2 US-08-443-260-3	Sequence 3, Appl
41	97	5.9	2955	3 US-08-442-805A-3	Sequence 3, Appl
42	97	5.9	2955	3 US-08-443-900A-3	Sequence 3, Appl
43	97	5.9	2955	4 US-08-444-818-124	Sequence 124, App
44	97	5.9	2955	4 US-08-249-843-3	Sequence 3, Appl
45	97	5.9	3011	2 US-08-833-678A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-789-078-1
; Sequence 1, Application US/08789078
; Patent No. 5843885
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LEA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,078
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,513
; FILING DATE: 19-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: (816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:

TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-476-062A-45

Query Match	6.6%;	Score 109.5;	DB 2;	Length 769;
Best Local Similarity	22.3%;	Pred. No. 0.029;		
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;				
QY	1	MVTQLKGGTQMCLICDNRSTRELKALHLVPOYQDNNYLQSDVPKPMTALVGRLFPASTK	60	
		: : : : : : : : : : : : : : : : : : : :		
Db	436	IYTVIVLPQCECRQSRDRSL--CH-----GKFLECGTCRCDTGVIYNKECQTQ	486	
QY	61	LNLITOOLEGA-----LPSVWNGSAFPGSGTLPGPKKITLAGYDCGFA-----	103	
		: : : : : : : : : : : : : : : : : : : :		
Db	487	-GRSSQELFGSCRKDNNSIISGLGCVCQCCLHTSDVPG--KLIIYGQYCECDTINCR	543	
QY	104	-SGDFCNN-----CNCNCCNNLLHHIDIEF-----KAIKACLGNPEAFPTKGKQLGN	152	
		: : : : : : : : : : : : : : : : : : : :		
Db	544	YNQVCGGPGRGUFCGKC--RCHFEGSACOCERTTECL--NPRRVCS-GRGR---	595	
QY	153	VKPQHKNGCNRRSGCLKNYCEYEQAIM-----CSSIC----KCIQCKNYEESP	198	
		: : : : : : : : : : : : : : : : : : : :		
Db	596	-----CRC-----NVCECHSGVOLPLCOBCPCSPCGRYTSCACLFKEKFP	638	

4
RESULT
US-07-728-215-31
: Sequence 31, Application US/07728215
: Patent No. 5962643
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Dean
: APPLICANT: Quaranta, Vito
: APPLICANT: Pytela, Robert
: TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
: TITLE OF INVENTION: Thereof
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: United States of America
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/07728, 215
: FILING DATE: 19910711
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: F31 8717
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 31:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 769 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single

TOPOLOGY: linear
US-07-728-215-31

Query Match 6.6%; Score 109.5; DB 2; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVTICQLKGGTOMLCIDNSRTELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
Db 436 IVTVQVLPOCECRCDQSRDSL--CH-----GKFLGECICRCDTGYIGNKCECQTQ 486
QY 61 LNLITQOLEGA-----LPSVWNGSAPFSGSTLPGPKITLAGYCDGCA-----103
Db 487 -GRSSQLEGGSCRRKDNNSIICSLGDCVCGQCLCHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCNN-----CNCNCCNLLHHDIERF-----KAIKACLGRLNPEAFQPKIGKGLGN 152
Db 544 YNGQVCGGPRGLCFGCK--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR---595
QY 153 VKPOHNKGNCRSGCLKNYCEYEAQIM-----CSSIC-----KCIGCKNVEESP 198
Db 596 -----CRC-----NVCECHSGYQLPQCPCPSPCGKYISCAECLAFKGP 638

RESULT 5
US-08-938-085A-31
Sequence 31, Application US/08938085A
Patent No. 6339148
GENERAL INFORMATION:
APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,085A
FILING DATE: 26-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,215
FILING DATE: 11-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parent, Annette S.
REGISTRATION NUMBER: 42,058
REFERENCE/DOCKET NUMBER: 023070-080210US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-085A-31

Query Match 6.6%; Score 109.5; DB 4; Length 769;

Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVTICQLKGGTOMLCIDNSRTELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
Db 436 IVTVQVLPOCECRCDQSRDSL--CH-----GKFLGECICRCDTGYIGNKCECQTQ 486
QY 61 LNLITQOLEGA-----LPSVWNGSAPFSGSTLPGPKITLAGYCDGCA-----103
Db 487 -GRSSQLEGGSCRRKDNNSIICSLGDCVCGQCLCHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCNN-----CNCNCCNLLHHDIERF-----KAIKACLGRLNPEAFQPKIGKGLGN 152
Db 544 YNGQVCGGPRGLCFGCK--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR---595
QY 153 VKPOHNKGNCRSGCLKNYCEYEAQIM-----CSSIC-----KCIGCKNVEESP 198
Db 596 -----CRC-----NVCECHSGYQLPQCPCPSPCGKYISCAECLAFKGP 638

RESULT 6
PCT-US95-04886-1
Sequence 1, Application PC/TUS9504886
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELEPHONE: (816) 474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496

OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540
OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label- repeat
OTHER INFORMATION: /note- "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label- trans
OTHER INFORMATION: /note- "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label- cyto
OTHER INFORMATION: /note- "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
AUTHORS: Power,
TITLE: LFA-1 Amino acid sequence (B2) (from human
TITLE: tonsil)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
PCT-US95-04886-1

Query Match 6.6%; Score 109.5; DB 5; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
Qy 1 MVICOLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLSQSDVPKPMALVGRFLPASTK 60
Db 436 IVTVQVLPQCECRCDQSRDSL--CH-----GKGFLECGICRCDTG YIGKNECQTQ 486
Qy 61 LNLITQOLEGA-----LPSVNGSAFPSTLPGPKITLAGYCDCA----- 103
Db 487 -GRSQELEGSCRKDNNSIICSLGDCVCGQCLCHTSDVPG--KLIYGYCECDTINCER 543
Qy 104 -SGDFCENN-----CNCNCCNNLHDIERF-----KAKACILGRNPEAFQPKIGKGLGN 152
Db 544 YNGVCGGPGRLCGKGC--RCHPGFEGSACQCRTEGCL--NPRRVECS-GRGR--- 595
Qy 153 VKPQHKGCNRRSGCLKNYCECYEAQIM-----CSSIC-----KIGCKNYEESP 198
Db 596 -----CRC-----NVCEHSGYQLPLCQECPCGCPSPGKYISCAECLKFEKGP 638

RESULT 7
PCT-US96-01314-45
Sequence 45, Application PC/TUS9601314
GENERAL INFORMATION:
APPLICANT: M. Amin Arnaout
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
TITLE OF INVENTION: ANTAGONISTS
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts

COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01314
FILING DATE: 30-JAN-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/380,167
FILING DATE: 30-JAN-95
ATTORNEY/AGENT INFORMATION:
NAME: John W. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/267001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 769
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US96-01314-45
Query Match 6.6%; Score 109.5; DB 5; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
Qy 1 MVICOLKGGTQMLCIDNSRTRELKALHLVPOYQDNNYLSQSDVPKPMALVGRFLPASTK 60
Db 436 IVTVQVLPQCECRCDQSRDSL--CH-----GKGFLECGICRCDTG YIGKNECQTQ 486
Qy 61 LNLITQOLEGA-----LPSVNGSAFPSTLPGPKITLAGYCDCA----- 103
Db 487 -GRSQELEGSCRKDNNSIICSLGDCVCGQCLCHTSDVPG--KLIYGYCECDTINCER 543
Qy 104 -SGDFCENN-----CNCNCCNNLHDIERF-----KAKACILGRNPEAFQPKIGKGLGN 152
Db 544 YNGVCGGPGRLCGKGC--RCHPGFEGSACQCRTEGCL--NPRRVECS-GRGR--- 595
Qy 153 VKPQHKGCNRRSGCLKNYCECYEAQIM-----CSSIC-----KIGCKNYEESP 198
Db 596 -----CRC-----NVCEHSGYQLPLCQECPCGCPSPGKYISCAECLKFEKGP 638
RESULT 8
US-08-630-172-10
Sequence 10, Application US/08630172
Patent No. 6060054
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/630,172
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 676 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-630-172-10

```

Query Match	6.6%;	Score 108;	DB 3;	Length 676;	
Best Local Similarity	21.5%;	Pred. No. 0.034;			
Matches	51;	Conservative 26;	Mismatches 86;	Indels 74;	Gaps 14;
QY	1	MVTCOLKGGTQMLCIDNSRTRALKALHLVPOYQDNVNYLQSDVPKPMTALVGRFLPASTK	60		
DB	414	IVTVOVLPOCECRCDQSDRSL--CH-----GKGLEGGICRCDTQYIGKNCECQTQ	464		
QY	61	LNLTIOLEGA-----LPSVVNGSAFPGSGTLPGPKKITIAGYCDFA-----	103		
DB	465	-GRSSQLELGSCRKDNNSIICGLGDCVCGQLCHTSDVPQ--KLIIQYQYCECDTINCR	521		
QY	104	-SGDFCNN-----CNCNCCNLLHHDTERF-----KAIKACLGKNPFAQFIKGQGLGN	152		
DB	522	YNGQVCGGPGRLGFCGKC--RCHPGFGSCQCERTEGCL--NPRVECS-GRGR---	573		
QY	153	VKPOHNKGCNRRSGCLKNYCECYEAQIM-----CSSTKCIKGCKNYEESP	198		
DB	574	-----CRQ-----NVCECHSGYQLPLQCBQCPGCSGKGIYISACELKFFKPG	615		

RESULT 9
US-09-375-419-10
; Sequence 10, Application US/09375419
; Patent No. 6264950
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYPHOCYTE VETO
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/375,419
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/630,172
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 676 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-375-419-10

```

Query Match 6.6%; Score 108; DB 4; Length 676;
Best Local Similarity 21.5%; Pred. No. 0.034;
Matches 51; Conservative 26; Mismatches 86; Indels 74; Gaps 14;

Qy	1	MVICQLKGGTOML	CINSTR	RELKALHLV	PQYQDQNN	YLOS	DPKDPMT	ALVGRFL	PASTK	60
		:	:	:	:	:	:	:	:	:
Db	414	IVTVQVL	PQCECR	QDRSL	-CH-	-----	GRGFL	EGCIG	CDTQY	IGKNCE
		:	:	:	:	:	:	:	:	:
Qy	61	LNLI	TOOLEGA	-----	LP	SVANG	APFS	SGTLP	GGPKIT	LACYDC
		:	:	:	:	:	:	:	:	:
Db	465	-GRSS	QLEGS	CRKDN	NSI	ICSLG	DCVCG	QC	LCHT	SDVFG
		:	:	:	:	:	:	:	:	:
Qy	104	-SGD	FCNN	-----	CNC	NNC	NNL	HH	DIERF	-----
		:	:	:	:	:	:	:	:	:
Db	522	YNGQ	VGCG	PGRL	CF	CGKC	-RCH	PG	FE	GSAC
		:	:	:	:	:	:	:	:	:
Qy	153	VKPH	NKNC	RRSG	CL	KNY	CEY	AOIM	-----	CSSIC
		:	:	:	:	:	:	:	:	:
Db	574	-----	CR	-----	NV	CE	CHSG	YQLP	LC	QCE
		:	:	:	:	:	:	:	:	:

```

RESULT 10
US-08-444-818-148
; Sequence 148, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

US-08-444-818-148

Query Match 6.6%; Score 108; DB 4; Length 739;
Best Local Similarity 24.9%; Pred. No. 0.038; 79; Indels 70; Gaps 12;
Matches 56; Conservative 20; Mismatches 20

QY 79 SAPP-SGSTLPKPPKITTLAGYDCFCASGDFCNCCNCCNNL-----HHDIERFKAI 130
Db 233 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNTCTQTVDFSLDPTFTIETILP 287
QY 131 KACLGRNPEAFQPKIGKQGLGNVQHNK--GCNRRSGCLN--YCEYEQAQIMCSSIC 186
Db 288 QDAVSRTOR-----RGRTGRKPGINRFVAPGERPSGMFSSVLCYDA-----332
QY 187 KCIGCKNYEESPERKTLMSMPNVMOTGGL-----EG-----SHYLPPTKFS 227
Db 333 ---GCAYELTP-AETTVRLRAYMTPGLPVQODHLEFWEVFTGLTHDAHFLSOTKOS 388
QY 228 G--LPRFSDRRPSSCISWEVVEATCACLQAQEEAEKEHCSKCL 270
Db 389 GENLP-----YLVAYQATVCARAQAPPSPSDQMWKCL 420

RESULT 11

US-08-444-818-138
; Sequence 138, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 138:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2995 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-138

Query Match 6.6%; Score 108; DB 4; Length 2995;
Best Local Similarity 24.9%; Pred. No. 0.23;
Matches 56; Conservative 20; Mismatches 79; Indels 70; Gaps 12;

QY 79 SAPP-SGSTLPKPPKITTLAGYDCFCASGDFCNCCNCCNNL-----HHDIERFKAI 130

Db 1424 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNTCTQTVDFSLDPTFTIETILP 1478
QY 131 KACLGRNPEAFQPKIGKQGLGNVQHNK--GCNRRSGCLN--YCEYEQAQIMCSSIC 186
Db 1479 QDAVSRTOR-----RGRTGRKPGINRFVAPGERPSGMFSSVLCYDA-----1523
QY 187 KCIGCKNYEESPERKTLMSMPNVMOTGGL-----EG-----SHYLPPTKFS 227
Db 1524 ---GCAYELTP-AETTVRLRAYMTPGLPVQODHLEFWEVFTGLTHDAHFLSOTKOS 1579
QY 228 G--LPRFSDRRPSSCISWEVVEATCACLQAQEEAEKEHCSKCL 270
Db 1580 GENLP-----YLVAYQATVCARAQAPPSPSDQMWKCL 1611

RESULT 12

US-08-833-678A-2
; Sequence 2, Application US/08833678A
; Patent No. 5989905
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,678A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/529,169
; FILING DATE: 15-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-833-678A-2

Query Match 6.1%; Score 101; DB 2; Length 465;
Best Local Similarity 24.8%; Pred. No. 0.099;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAPP-SGSTLPKPPKITTLAGYDCFCASGDFCNCCNCCNNL-----HHDIERFKAI 130
Db 232 SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNTCTQTVDFSLDPTFTIETILP 286

```

332 ----GCAWYELTP-AETTVRLRAYMNTPGLPVQDHLFEWGVTGLTHDAHFLSQTK 386
Db

227 SG--LPRFSDRRPPSCISWEVVEATCACILLAGEEAEKEHCCKL 270
QY      |||          :   |   |||
        :             :   |   |||

387 SGENLP-----YLVAOATVCARQAQPSPDDWMKCL 419
Db

```

RESULT 14

US-08-833-678A-1
Sequence 1, Application US/08833678A
Patent No. 5989905
GENERAL INFORMATION:
APPLICANT: HOUGHTON, MICHAEL
APPLICANT: CHOO, QUL-LIM
APPLICANT: HAN, JANG
APPLICANT: CHOE, JOONHO
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/833,678A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/529,169
FILING DATE: 15-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0100.005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3274
TELEFAX: (510) 655-3542
TELEX: n/a
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-833-678A-1

Query Match 6.1%; Score 101; DB 2; Length 631;
Best Local Similarity 24.8%; Pred. No. 0.15;
Matches 56: Conservative 21: Mismatches 77: Indels 72: Gaps 13:

Qy	79	SAPP-SGSTLPGPPKITLAGYDCFASGDPCNNCNCNCCNL-----HHDIERFKAI	130
		: : : : : :	
Db	398	SVIPTS GDVVVATDALMTGY-----TGDFSDVIDCNTCTQTDFSLDPTFTTETILP	452
Qy	131	KACLG RNP EAFQPIKGGLG---NVKPQH NKCNCRRSC LKN--YCCEYEAQIMCSSI	185
		: : : : : : : : : :	
Db	453	QDAVSRTQR--RGRTGRGPGIYRFAPGE-----RPSGMFDSSVLCECYDA-----	497
Qy	186	CKCI GCKNYEESPERKTLMSPNPMOTGGU-----EG-----SHYLPTTKF	226
		: : : :	
Db	498	----GCAYWELTP-AETTIVRLRAYMNTPLGPVCQDHLEFPWEVFGTGTHIDAHFLSQTKQ	552
Qy	227	SG---LPFRSHDRPSSCSISWEVYEATCACLLAAGEEAEKECHSKCL	270

RESULTS 13

US-08-529-169A-2
Sequence 2, Application US/08529169A
Patent No. 6194140
GENERAL INFORMATION:
APPLICANT: HOUGHTON, MICHAEL
APPLICANT: CHOO, QUL-LIM
APPLICANT: HAN, JANG
APPLICANT: CHOE, JOONHO
TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
HELICASE ACTIVITY AND IMPROVED SOLUBILITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: Intellectual Property - R440, P.O. Box 8097
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/529,169A
FILING DATE: 15-SEP-1995
CLASSIFICATION: 4325
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0100.005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-3274
TELEFAX: (510) 655-3542
TELEX: n/a
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-529-169A-2

Query Match	6.1%	Score 101;	DB 4;	Length 465;
Best Local Similarity	24.8%	Pred. No. 0.099;		
Matches 56: Conservative	21;	Mismatches 77;	Indels 72;	Gaps 13;

[illegible]

us-09-743-237-5.rai

Fri Jul 12 08:35:30 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:20:36 ; Search time 210.79 seconds
(without alignments)
499.275 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDSRT.....GRCLSOILHTEFKSKGLKWE 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_Main:*

- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pap.*
- 2: /cgn2_6/ptodata/2/paa/US06_COMB.pap.*
- 3: /cgn2_6/ptodata/2/paa/US07_COMB.pap.*
- 4: /cgn2_6/ptodata/2/paa/US08_COMB.pap.*
- 5: /cgn2_6/ptodata/2/paa/US081_COMB.pap.*
- 6: /cgn2_6/ptodata/2/paa/US082_COMB.pap.*
- 7: /cgn2_6/ptodata/2/paa/US083_COMB.pap.*
- 8: /cgn2_6/ptodata/2/paa/US084_COMB.pap.*
- 9: /cgn2_6/ptodata/2/paa/US085_COMB.pap.*
- 10: /cgn2_6/ptodata/2/paa/US086_COMB.pap.*
- 11: /cgn2_6/ptodata/2/paa/US087_COMB.pap.*
- 12: /cgn2_6/ptodata/2/paa/US088_COMB.pap.*
- 13: /cgn2_6/ptodata/2/paa/US089_COMB.pap.*
- 14: /cgn2_6/ptodata/2/paa/US090_COMB.pap.*
- 15: /cgn2_6/ptodata/2/paa/US091_COMB.pap.*
- 16: /cgn2_6/ptodata/2/paa/US092_COMB.pap.*
- 17: /cgn2_6/ptodata/2/paa/US093_COMB.pap.*
- 18: /cgn2_6/ptodata/2/paa/US094_COMB.pap.*
- 19: /cgn2_6/ptodata/2/paa/US095_COMB.pap.*
- 20: /cgn2_6/ptodata/2/paa/US096_COMB.pap.*
- 21: /cgn2_6/ptodata/2/paa/US097_COMB.pap.*
- 22: /cgn2_6/ptodata/2/paa/US098_COMB.pap.*
- 23: /cgn2_6/ptodata/2/paa/US099_COMB.pap.*
- 24: /cgn2_6/ptodata/2/paa/US100_COMB.pap.*
- 25: /cgn2_6/ptodata/2/paa/US101_COMB.pap.*
- 26: /cgn2_6/ptodata/2/paa/US60_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	100.0	299	21	US-09-743-237-5
2	1251	76.0	295	21	US-09-743-237-4
3	1251	76.0	295	21	US-09-743-237-24
4	533.5	32.4	223	1	PCT-US01-08656-10693
5	470	28.5	403	26	US-60-161-932-1700
6	470	28.5	890	26	US-60-167-217-12948
7	470	28.5	950	20	US-09-614-150-12897

RESULT 1

US-09-743-237-5
; Sequence 5, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YUJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 5
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-743-237-5

ALIGNMENTS

8	470	28.5	950	26	US-60-191-637-12935	Sequence 12935, A
9	446	27.1	147	1	PCT-US01-01329-1409	Sequence 1409, Ap
10	446	27.1	147	1	PCT-US01-01339-3988	Sequence 3988, Ap
11	446	27.1	147	21	US-09-764-891-3988	Sequence 3988, Ap
12	446	27.1	147	24	US-10-080-090-1409	Sequence 1409, Ap
13	437	26.5	438	14	US-09-087-136-13	Sequence 13, Appl
14	437	26.5	438	14	US-09-087-136-13	Sequence 13, Appl
15	437	26.5	438	16	US-09-220-091-13	Sequence 13, Appl
16	378	23.0	280	1	PCT-US01-08631-48317	Sequence 48317, A
17	376	22.8	524	21	US-09-708-427-33367	Sequence 33367, A
18	376	22.8	550	21	US-09-708-427-33366	Sequence 33366, A
19	376	22.8	571	21	US-09-708-427-33365	Sequence 33365, A
20	369	22.4	615	21	US-09-733-089-22002	Sequence 22002, A
21	369	22.4	615	21	US-09-733-089-22183	Sequence 22183, A
22	369	22.4	615	21	US-09-733-089-22184	Sequence 22184, A
23	369	22.4	615	22	US-09-816-660-22002	Sequence 22002, A
24	369	22.4	615	22	US-09-816-660-22183	Sequence 22183, A
25	369	22.4	615	22	US-09-816-660-22184	Sequence 22184, A
26	366.5	22.3	241	26	US-60-324-109-27382	Sequence 27382, A
27	357	21.7	603	19	US-09-573-655A-2154	Sequence 2154, Ap
28	357	21.7	603	19	US-09-573-655A-2152	Sequence 2152, Ap
29	357	21.7	603	21	US-09-708-427-27294	Sequence 27294, A
30	353	21.4	534	21	US-09-708-427-27295	Sequence 27295, A
31	351	21.3	518	21	US-09-708-427-27296	Sequence 27296, A
32	316.5	19.2	207	21	US-09-733-089-22013	Sequence 22013, A
33	316.5	19.2	207	22	US-09-816-660-22013	Sequence 22013, A
34	314.5	19.1	168	21	US-09-733-089-21980	Sequence 21980, A
35	314.5	19.1	168	22	US-09-816-660-21980	Sequence 21980, A
36	278	16.9	53	1	PCT-US00-26524B-6922	Sequence 6922, Ap
37	263	16.0	349	18	US-09-428-944-1881	Sequence 1881, Ap
38	257.5	15.6	386	26	US-60-324-109-16915	Sequence 16915, A
39	257.5	15.6	556	26	US-60-324-109-32886	Sequence 32886, A
40	245.5	14.9	497	21	US-09-708-427-25631	Sequence 25631, A
41	245.5	14.9	497	21	US-09-708-427-25630	Sequence 25630, A
42	245.5	14.9	658	21	US-09-708-427-25629	Sequence 25629, A
43	245	14.9	770	26	US-60-324-109-17053	Sequence 17053, A
44	245	14.9	770	26	US-60-324-109-17063	Sequence 17063, A
45	242.5	14.7	402	26	US-60-324-109-31233	Sequence 31233, A

Query Match 100.0%; Score 1647; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 6.5e-148;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVICQLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYQSDVPKPMVLRGRLPASTK 60
DB 1 MVICQLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYQSDVPKPMVLRGRLPASTK 60


```

RESULT      11
US-09-764-891-3988
; Sequence 3988, Application US/09764891
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3988
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3988

```

[illegible]

```

RESULT 12
US-10-080-090-1409
; Sequence 1409, Application US/10080090
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA120C1
; CURRENT APPLICATION NUMBER: US/10/080,090
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 3506
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1409
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-080-090-1409

```

	Query Match	27.1%	Score 446;	DB 24;	Length 147;
	Best Local Similarity	98.9%;	Pred. No.	6.9e-34;	
	Matches 87; Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;
Qy	1	MVICQLKGGTQMCLIDNSRTELKALHLVPQYQQNNVYLQSDVKPMTALVGRELPASTK	60		
Dd	60	MVICQLKGGTQMCLIDNSRTELKALHLVPQYQQNNVYLQSDVKPMTALVGRELPASTK	119		
Qy	61	LNLITQOLEGALSVMNGSAFPGSGTLTP	88		
Dd	120	LNLITQOLEGALSVMNGSAFPGSGTLTP	147		

```

RESULT 13
US-09-087-136-13
; Sequence 13, Application US/09087136
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PR1
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

```

	Query Match	26.58;	Score 437;	DB 14;	Length 438;
	Best Local Similarity	41.3%;	Pred. No.	2.2e-32;	
Matches	92; Conservative	35;	Mismatches	62;	Indels 34; Gaps
Qy	98 YCDCFASGDFCNCCNCCNNLHHDIERKAIACILGRNPFAFPKTG--KGOLGNVKP	155			
	: : :				
Dd	191 YCDCFANGFCRDCNCXDKCHNI EYDSORSRAIQSLERNPNAPFKIGIARGGITDIER	250			
	: : :				
Qy	156 QHNKGCNRSGCLUKNYCECYEAQTMCSSICKTCCKN-----YEES-----PERKLT	203			
	: : :				
Dd	251 LHQGCGCHKSGCLKNYCECYEAVPCPTDRCKCKGCQNTETRYMTRYKNSGGAVSNTNAL	310			
	: : :				
Qy	204 MSMPNYMQT-----GGGL----EGSHVLPPTKPSGLPFRESHDRR--PSSCISMEWVEAT	250			
	: : :				
Dd	311 MSLTNASTATPDSPGSVSVTDEHGDDY-EDMLLSHRPKVEMDPRRFFPWYYMTDEVVEAA	369			
	: : :				
Qy	251 CACLQAAGEA-----EKHCSCKLAEQMILEBFGRLCSQL	287			
	: : :				
Dd	370 TMCVHAQEALNKYEQVOTDEKLIINNEKLVLRBFPGRCLEOMI	412			
	: : :				

```

RESULT 14
US-09-087-136-13
; Sequence 13, Application US/09087136A
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136A
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

```

```
Query Match          26.5% Score 437; DB 14; Length 438;
Best Local Similarity 41.3%; Pred. No. 2.2e-32;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;
```

Qy 98 YCDCFASGDFCNCCNNCCNHLHHDIERFKAIKACLRGNPEARQPKTG--KGQLGNVKP 155
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 191 YCDCFANGFCRDCNCCKDCHNIEYDSQRSAIKQSLERNPNAPKPGICARGGITDIER 250

Qy 156 QHNKGNCRRSGCLKNTCCEYEAQIMCSSICKIGCKN-----YEES-----PERKTL 203
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 251 LHQGCCHKKSGLCKNTCECYEAKVPTCDRCRGCONTEYRMTRYKNSGGAVSNTNAL 310

Search completed: July 11, 2002, 08:20:36
Job time: 420 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:14:15 ; Search time 30.19 seconds
(without alignments)
1028.575 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGQTMLCIDNSRT.....GRCLSQILHTFKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 323636 seqs, 103855142 residues

Total number of hits satisfying chosen parameters: 323636

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	366.5	22.3	282	6	US-10-155-881-9626
2	357	21.7	603	5	US-09-573-655B-2154
3	357	21.7	603	5	US-09-573-655B-2192
4	348	21.1	510	6	US-10-155-881-9875
5	278	16.9	53	6	US-10-106-698-6932
6	248.5	15.1	896	6	US-10-135-881-20031
7	201.5	12.2	359	6	US-10-155-881-9183
8	183	11.1	143	6	US-10-155-881-9877
9	167	10.1	120	6	US-10-138-145-1127
10	112	6.8	1774	1	PCT-US02-01339-6
11	109.5	6.6	699	1	PCT-US02-09671-718
12	109.5	6.6	761	1	PCT-US02-09671-711
13	109.5	6.6	769	1	PCT-US02-09671-708
14	109.5	6.6	769	1	PCT-US02-09671-710
15	109.5	6.6	769	1	PCT-US02-09671-712
16	109.5	6.6	769	1	PCT-US02-09671-713
17	109.5	6.6	769	1	PCT-US02-09671-714
18	109.5	6.6	769	1	PCT-US02-09671-715
19	109.5	6.6	769	1	PCT-US02-09671-716
20	109.5	6.6	769	1	PCT-US02-09671-719
21	109.5	6.6	769	1	PCT-US02-09671-720
22	109.5	6.6	769	6	US-10-072-844-31
23	109.5	6.6	769	6	US-10-072-838-31
24	109.5	6.6	769	6	US-10-099-007A-12
25	109.5	6.6	817	6	US-10-144-793-141
26	108	6.6	676	1	PCT-US02-09671-717

27	106	6.4	200	6	US-10-155-881-9182	Sequence 9182, Ap
28	105.5	6.4	2180	7	US-60-360-039-5009	Sequence 5009, Ap
29	100	6.1	1986	6	US-10-123-962-2	Sequence 2, Appl
30	100	6.1	2017	6	US-10-123-962-4	Sequence 4, Appl
31	99	6.0	2195	7	US-60-360-039-6828	Sequence 6828, Ap
32	98.5	6.0	556	1	PCT-US02-09944-486	Sequence 486, App
33	98.5	6.0	574	7	US-60-360-039-5730	Sequence 5730, Ap
34	98.5	6.0	2749	6	US-10-123-155-385	Sequence 385, App
35	98.5	6.0	2749	6	US-10-137-871-385	Sequence 385, App
36	98.5	6.0	2749	6	US-10-141-761-385	Sequence 385, App
37	98.5	6.0	2749	6	US-10-140-864-385	Sequence 385, App
38	98.5	6.0	2749	6	US-10-140-923-385	Sequence 385, App
39	98.5	6.0	2749	6	US-10-141-756-385	Sequence 385, App
40	98.5	6.0	2749	6	US-10-141-759-385	Sequence 385, App
41	98.5	6.0	2749	6	US-10-140-472-385	Sequence 385, App
42	98.5	6.0	2749	6	US-10-140-805-385	Sequence 385, App
43	98.5	6.0	2749	6	US-10-142-885-385	Sequence 385, App
44	98.5	6.0	2749	6	US-10-146-731-385	Sequence 385, App
45	98.5	6.0	2749	6	US-10-142-426-385	Sequence 385, App

ALIGNMENTS

RESULT 1

US-10-155-881-9626
; Sequence 9626, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9626
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9626

Query Match 22.3%; Score 366.5; DB 6; Length 282;

Best Local Similarity 47.7%; Pred. NO. 1.3e-23;

Matches 63; Conservative 22; Mismatches 26; Indels 21; Gaps 3;

Qy 98 YDCDFASGDFCNCNCNCCNHLHDIERFKAICACLRNPEAFOPKICKGOL---GNVK 154

Db 82 YCECFASGTGDCGCTCFNPNDEVARREAVEAILDRNPDAPFRKIGSSPLMHRNEV 141

Qy 155 P-----QHNGKGNCRSGCLKNYCEYCAQIMCSCSSICKICGKNYBESPERKTLMSMP 207

Db 142 PSDLPVGRKHNGCHCKSGCLKYCECFQANILSENCKMCKDFEGSEERNL---- 197

Qy 208 NYMTGGLEGSH 219

Db 198 -----PQGDH 202

RESULT 2

US-09-573-655B-2154
; Sequence 2154, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptid
; FILE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18

```

; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2154
; LENGTH: 603
; TYPE: prt
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-2154

```

Query Match 21.7%; Score 357; DB 5; Length 603;
Best Local Similarity 28.2%; Pred. No. 2e-22;
Matches 96; Conservative 43; Mismatches 83; Indels 118; Gaps 12;

Qy	39	LQSDVP	---KP-MTALVGRF----	LPASTKLNLIQTQLEGALPSVNGSAFFPSGSTLP	88
Db	58	ISSVSPSTIRPGMTIAIGQVTVQVRPLTPMATMNSPPSQQ-----	IVNA---PTRHP	109	
Qy	89	GPPKI	-----TLAGYCDCFASGD	FNNCNCNCCNLLHH	122
Db	110	ESPKARGPRPVEGRDGTPOKKQKCCKHSRCLKLYCECFASGT	YCDGCVNCFNNVDN	169	
Qy	123	DIERFAIKACLRNEAPQPKICKQOLN-----	VKPOHKGCCNRRSCCLKN	171	
Db	170	EPARREAVEATLRNPFAPRPKTA	SSPHGGDRKREDIGEVLGKHKGCHCKKGCCLKK	229	
Qy	172	YCCEYEAQIMCSCKICGICKYEE	SPERKTLMSPN-----YMQ-----	TGGLEGSH	219
Db	230	YCECFQANIILCSNCKCLDCKNFEGSEERQALPHGESHNMAY	LQOANAAITGAVSSG	289	
Qy	220	YLP-----	PTFESGLPRSHDR--PSS-----	240	
Db	290	FAPSPAPRRKKGIELFNQAIKDSRLSHFPQVNNRGTTGPTSGT	SPSPVSRAGGNASSV	349	
Qy	241	-----	CISVEVVEATCACLQAGEAEKECHSK	268	
Db	350	PSKFVYBSLIADITOPHDVRAICSVLTAVAGEAAKTSTDK	389		

```

RESULT      3
US-09-573-655B-2192
; Sequence 2192, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVIEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2192
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-2192

```

Query Match 21.7%; Score 357; DB 5; Length 603;
Best Local Similarity 28.2%; Pred. No. 2e-22;
Matches 96: Conservative 43; Mismatches 83; Indels 118; Gaps 12;

[illegible]

Qy	172	YCECYEAQIMCSCICKGCKNYEESPRTKLMSFN----	YMO-----TGGLEGSH	219
		: :	: :	
Db	230	YCEFQANILCSNCKLDCNKPEEGEEKRALPHGHSHMAYLQAANAATGVAGSSG	289	
		: :	: :	
Qy	220	YLP-----PTKFSGLPRESHDDR--PSS-	240	
		: :	: :	
Db	290	FAPSPAPRRKGQEILFNQAIKDSSRLSHFPVQNNNGRTGGTSTPSPVSRAGGNASSV	349	
Qy	241	-----CISVEVVEATCACLLAQGEAEKECHSK	268	
		: : :	: : :	
b	350	pskevvsbslIADITOPHDVPALICSVLTVTVAGEAAKTSTDK	389	
		: : :	: : :	

RESULT

```

US-10-155-881-9875
; Sequence 9875, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutflyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9875
; LENGTH: 510
; TYPE: PRT
; ORGANISM: zea mays
US-10-155-881-9875

```

Query Match 21.1%; Score 348; DB 6; Length 510;
Best Local Similarity 32.6%; Pred. NO. 9.5e-22;
Matches 79; Conservative 26; Mismatches 57; Indels

Qy	66	QQLGALPSVVNGSAF-----PSGSTL-PCPPKKITLAG-----	97
		: : : : : : :	
Db	17	QOMEPPALVREAGAEFRAPROTQOQNPNTAVQVQSPVLRPWPMBEITLSMKPVVEMKS	76
Qy	98	-----YCDGCFASGDFFCNCCNCCNCCNLLHHDTERFKAKACILGRN	137
		: : : : : : :	
Db	77	GTPAKKKHCKNKSQCLKLYCEGFAAGDYCDGCNKCQCGTVENEGROEAINNTKLRN	136
Qy	138	PEAFQPKIGKQGLN-----VFPQHNKGCNCRSSGCLKNKYCEYEAQIMCISICK	187
		: : : : : : : :	
Db	137	PNATQPKTIENGPIPSVVMKDAGALPSHPKHNGCHCKKSGCLKYCYCFQANILCSKNCK	196
Qy	188	CIGCKNYEESPE-----RKTLMSPNYMQTGGLEGSHYLPPTKFSGLPRPSHD	235
		: : : : : :	
Db	197	CMDCKNYEGSEELRSTTQGDNSCDNRNIQQANVALNGAIGSSGY-----RSPV	246
Qy	236	RR	237
Db	247	RR	248

5 RESULT

```

RESULT
US-10-106-698-6932
; Sequence 6932, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA003P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29

```

; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6932

Query Match 16.9%; Score 278; DB 6; Length 53;
Best Local Similarity 100.0%; Pred. No. 6.3e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 VEATCACLAAQEEAEKEHCKSLAEQMIIEFGRCLSQILHTEFKSKGLKME 299
|||||
Db 1 VEATCACLAAQEEAEKEHCKSLAEQMIIEFGRCLSQILHTEFKSKGLKME 53

RESULT 6

US-10-155-881-20031
; Sequence 20031, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 20031
; LENGTH: 896
; TYPE: PRT
; ORGANISM: Glycine max
US-10-155-881-20031

Query Match 15.1%; Score 248.5; DB 6; Length 896;
Best Local Similarity 40.7%; Pred. No. 5.3e-13;
Matches 50; Conservative 13; Mismatches 31; Indels 29; Gaps 5;

QY 98 YDCDFASGDFCANN-CNCCNCCNHHDIERFKAIAKCL-GRNPEAFQPKIGKQGLGNVKP 155
|||||
Db 487 YDCDFAAGTYCTDPCACQCGCLNRPEY-VETVYETKQOIESRNPFIAPAKI-----VQP 538
QY 156 -----QHNGCNCRRSGCKLKNYCEYEAQIMCSSTCKIGCKNYEE 196
|||
Db 539 TTDISHHDDENLTTPSSARHRRGCNCKRSMCLKKYCEYQANVGCGSCRCGCKNVHG 598
QY 197 SPE 199
|
Db 599 KXE 601

RESULT 7

US-10-155-881-9183
; Sequence 9183, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9183
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9183

Query Match 12.2%; Score 201.5; DB 6; Length 359;
Best Local Similarity 23.1%; Pred. No. 1.8e-09;
Matches 63; Conservative 35; Mismatches 102; Indels 73; Gaps 9;

QY 98 YDCDFASGDFCANN-CNCCNCCN-----NLHHD-----IERFKAIAKCLGRNPEAFQPKIG 146
|||
Db 80 YPCFSGGGYCSKCCQPCFKKEAFETVHTTRKVLSSRQKMSKINRRPEANTEPWE 139
QY 147 KQQLGNVKKPQHNGCNCRRSGCKLKNYCEYEAQIMCSSTCKIGCKNYEESPRTKTLMSM 206
|||
Db 140 DAHSSSTTPPKRGCKNCKSSCLKKYCDYQDGTGCSLFCRCDDCQNPFGKNEGIMADDS 199
QY 207 PNYMTGG-----LEGS-----HYP-----TK 225
|||
Db 200 KRYLYTGADLDHSEGEHDFVVERSPRLQSPISKESFHQTPPHLRASSRDAHVFPFAISO 259
QY 226 FSGLPREFH--DRRPPSSCIS-----WEVVE-----ATCACLAAQEEAEKE 264
|||
Db 260 WQALPRSHWCHSKNRSSDRAMDDSNANYKSNHDYQVPHKEDSYSSKCVQILNGMAELS 319
QY 265 HCSKLAEQMIIEFGRCLSQILHTEFKSKGLK 297
|||
Db 320 QVEKSVAPDVFLQPGNREIFVSLSGDVRAMWLK 352

RESULT 8

US-10-155-881-9877
; Sequence 9877, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9877
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9877

Query Match 11.1%; Score 183; DB 6; Length 143;
Best Local Similarity 40.9%; Pred. No. 2.4e-08;
Matches 36; Conservative 16; Mismatches 24; Indels 12; Gaps 3;

QY 98 YDCDFASGDFCANNCCNCCNHHDIERFKAIAKCLGRNPEAFQPKIGK-QGLGNV--- 153
|||
Db 55 YCECFQELQYCDGCGNCSGNVGNENARNEAIEAIRQORNPSPAFQPKIGDNTLNVRKD 114
QY 154 -----XPQHNGCNCRRSGCKLKNYCE 174
|||
Db 115 KFWSSPSRSKPKHKSCHCKKS-CLKKYCD 141

RESULT 9

US-10-138-145-1127
; Sequence 1127, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen

APPLICANT: Shenk, Michael Andrew
APPLICANT: Gibson, John Bryan
APPLICANT: Norris, Michael Geoffrey
APPLICANT: Saulsbury, Keith Martin
APPLICANT: Hall, Claire
TITLE OF INVENTION: Compositions isolated from forage
FILE REFERENCE: 11000.10580
CURRENT APPLICATION NUMBER: US/10/138,145
CURRENT FILING DATE: 2002-05-01
NUMBER OF SEQ ID NOS: 1652
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1127
LENGTH: 120
TYPE: PRT
ORGANISM: Lolium perenne
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(120)
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-138-145-1127

Query Match 10.1%; Score 167; DB 6; Length 120;
Best Local Similarity 37.4%; Pred. No. 4.4e-07;
Matches 40; Conservative 15; Mismatches 36; Indels 16; Gaps 8;
QY 98 YDCDFASGDFCNN-CNCNCCNNLHDIERFKAICACLRNPEAFOPKI-----GKGQLGN 152
DB 16 YCEC-AAGYVCSEPCGCGCLNKPHETIV-LSTRQIEFRNPLA-APKVIRLSDAAOETQ 72
QY 153 VKP-----QHNKGNCRRSGCLNYCEYBAQIMCISICKICGCKN 193
DB 73 EDPNTPASARHKGNCCKKSSCLK-YCECYGGVGLWLTNCR-BCKN 117

RESULT 10
PCT-US02-01339-6
Sequence 6, Application PC/TUS0201339
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: LEE, Ernestine A.
APPLICANT: WALIA, Narinder K.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: AZIMZAI, Yalda
APPLICANT: TANG, Y. Tom
APPLICANT: YUE, Henry
APPLICANT: THANGAVELU, Kavitha
APPLICANT: XU, Yuming
APPLICANT: ARVIZU, Chandra
APPLICANT: WARREN, Bridget A.
APPLICANT: YAO, Monique G.
APPLICANT: AU-YOUNG, Janice
APPLICANT: HAFALIA, April J.A.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: KALLICK, Deborah A.
APPLICANT: GANDHI, Aneena R.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: KHAN, Farrah A.
APPLICANT: LU, Yan
APPLICANT: SWARNAKAR, Anita
APPLICANT: RAMKUMAR, Javalaxmi
APPLICANT: NGUYEN, Dannel B.
APPLICANT: GRAUL, Richard
APPLICANT: LU, Dyung Aina M.
TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
FILE REFERENCE: PI-0346 PCT
CURRENT APPLICATION NUMBER: PCT/US02/01339
CURRENT FILING DATE: 2002-01-16
PRIOR APPLICATION NUMBER: 60/265,927; 60/271,196; 60/274,549; 60/334,179
PRIOR FILING DATE: 2001-01-19; 2001-02-02; 2001-02-23; 2001-03-09; 2001-11-28
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PERL Program

SEQ ID NO 6
LENGTH: 1774
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7722591CD1
PCT-US02-01339-6

Query Match 6.8%; Score 112; DB 1; Length 1774;
Best Local Similarity 20.2%; Pred. No. 0.46;
Matches 51; Conservative 33; Mismatches 73; Indels 96; Gaps 15;
QY 86 TLPGPPIITLAGYCDCAFSGDFCNN---CN-CNNCC---NNLH-----H 122
DB 800 TLPDP-----CRLLSSPEACNOSGACTWCHGACLSGDAQHRLGCGGSPCSPMPSPE 852
QY 123 DIERFKAICACLRNPEAFOPKIGQLGNVYKPOHNKGCNCRSGCL-----KNYCE 174
DB 853 ECRRLRTCSECLARHPRTLQ--GDGEAST--PRCKWCTNCEGACIGRNGSCTSENDRCR 908
QY 175 CYEAQIMCSSIC-----KCIGCKNYEESPERKTLMSM-PNY----- 209
DB 909 INQREVEFWAGNCSEACGAADCEQCTREGKCMWTRQFKRTGETRRLSVQPTDWTFCFSH 968
QY 210 ---MOTGGLEGSHYLP-PTKFSGLPRFSDRRPSSCISWEVVEATCACLAAOGEAEKE 264
DB 969 SLLNVSPMPVSESPPLPCPTCHLLP-----NCTS-----CLDSKADGWQ 1010
QY 265 HC-----SKCLA 271
DB 1011 HCWSSSLQOCLS 1023

RESULT 11
PCT-US02-09671-718
Sequence 718, Application PC/TUS0209671
GENERAL INFORMATION:
APPLICANT: ZYCOS Inc.
TITLE OF INVENTION: TRANSLATIONAL PROFILING
FILE REFERENCE: 08191-026W01
CURRENT APPLICATION NUMBER: PCT/US02/09671
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: 60/279,495
PRIOR FILING DATE: 2001-03-28
PRIOR APPLICATION NUMBER: 60/292,544
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: 60/310,801
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/326,370
PRIOR FILING DATE: 2001-10-01
PRIOR APPLICATION NUMBER: 60/336,780
PRIOR FILING DATE: 2001-12-04
PRIOR APPLICATION NUMBER: 60/358,985
PRIOR FILING DATE: 2002-02-20
NUMBER OF SEQ ID NOS: 2041
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 718
LENGTH: 699
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-09671-718

Query Match 6.6%; Score 109.5; DB 1; Length 699;
Best Local Similarity 22.3%; Pred. No. 0.26;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVICQLKGGTQMLCIDNSRTRRELKALHLVPOYQDNVYVLOSVPMTALVGRFLPASTK 60
DB 436 IVTVQVLPQCECRCDOSDRSL--CH-----GKGFLECGICRCDTGYIGKNCECQTQ 486

Job time: 40 sec

Best Local Similarity 22.3%; pred. NO. 0.29;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

[illegible]

```

RESULT 15
PCT-US02-09671-712
, Sequence 712, Application PC/TUS0209671
, GENERAL INFORMATION:
, APPLICANT: Zycos Inc.
, TITLE OF INVENTION: TRANSLATIONAL PROFILING
, FILE REFERENCE: 08191-026W01
, CURRENT APPLICATION NUMBER: PCT/US02/09671
, CURRENT FILING DATE: 2002-03-28
, PRIOR APPLICATION NUMBER: 60/279,495
, PRIOR FILING DATE: 2001-03-28
, PRIOR APPLICATION NUMBER: 60/292,544
, PRIOR FILING DATE: 2001-05-21
, PRIOR APPLICATION NUMBER: 60/310,801
, PRIOR FILING DATE: 2001-08-08
, PRIOR APPLICATION NUMBER: 60/326,370
, PRIOR FILING DATE: 2001-10-01
, PRIOR APPLICATION NUMBER: 60/336,780
, PRIOR FILING DATE: 2001-12-04
, PRIOR APPLICATION NUMBER: 60/358,985
, PRIOR FILING DATE: 2002-02-20
, NUMBER OF SEQ ID NOS: 2041
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 712
, LENGTH: 769
, TYPE: prt
, ORGANISM: Homo sapiens
PCT-US02-09671-712

```

```
Query Match      6.8%; Score 109.5; DB 1; Length 769;  
Best Local Similarity   22.3%; Pred. No. 0.29;  
Matches    53; Conservative     26; Mismatches    84; Indels       75; Gaps        15;
```

QY	1	MVTICLKGGTQMCLCIDNSRTRKALHLVPQYDQNNYLQSDDVKKPMTALVGREFLPASTK	60
DG	:	: : :	:
DB	436	IYTVOVLPPQCRCRDOSDRSL--CH-----GKGFLCGICRCTGYIGKNCECOTQ	486
	:	: : :	:
QY	61	LNLITOLEGA-----LPSVVNVSAPFSGSTLPGPPIKITIAGYCDA----	103
	:	: :	:
DB	487	-GRSSOELEGSCRRKDNNIIISGIHGDCVCQGQLCHTSDPVG--KLIIYGQCYCEDTCINCR	543
	:	: :	:
QY	104	-SDGFECNN----CNCKNNCCNNLHHDTERE-----KAIKACLAGRNPEAFPKICKGGLGN	152
	:	: :	:
DB	544	YNGQVCGGPGRLGCFCGKC--RCHPGFEFGSNCQERTEGCL--NPRRVECS-GRGR---	595
	:	: :	:
QY	153	VKPQHNGKCNRSGCLKNTKYCEYEAQIM-----CSSTC-----KCIQKKNYEESP	198
	:	: :	:
DB	596	-----CRC-----NVCEHSIGYLPLCOBCPCGPCSPCGRYIVSAECKLFKEGP	638
	:	: :	:

Search completed: July 11, 2002, 08:14:16

Fri Jul 12 08:35:31 2002

us-09-743-237-5.rapn

Page 7

